



Corrigé du Devoir Surveillé 9

Variables aléatoires, intégration

Problème I - Variables aléatoires

Une urne (pour changer) contient $N \geq 2$ boules distinctes numérotées de 1 à N . On effectue successivement et avec remise des tirages dans l'urne. On fixe un espace probabilisé (Ω, \mathbb{P}) sur lequel on définit toutes nos variables aléatoires. Au tirage n , on note Z_n la variable aléatoire retournant 1 si le numéro tiré n'a jamais été pioché avant et 0 sinon. On note également Y_n le nombre de numéros **distincts** obtenus durant les n premiers tirages. On pose par convention $Y_0 = 0$ et on note que $Z_1 = Y_1 = 1$.

Partie 1 : Lois initiales

- On note que Z_2 ne possède que deux issues : 0 ou 1. Donc Z_2 suit une loi de Bernoulli. Déterminons son paramètre. Au deuxième tirage, un numéro a déjà été pioché et remis dans l'urne. Notons A_1 ce numéro. Dès lors, pour obtenir un nouveau numéro, il faut piocher une boule parmi les $N - 1$ qui n'ont pas le numéro A_1 : $N - 1$ choix alors que le nombre total de boules est de N (car on a remis la boule piochée au premier tirage). Le tirage étant uniforme sur toutes les boules :

$$\mathbb{P}(Z_2 = 1) = \frac{N - 1}{N}.$$

Conclusion,

$$Z_2 \sim \mathcal{B}\left(\frac{N - 1}{N}\right).$$

- On sait que lors du premier tirage on pioche forcément un nouveau numéro : $Y_1 = 1$. Puis lors du second tirage ou l'on repioche le même numéro, $Z_2 = 0$ et $Y_2 = 1$ ou l'on pioche un deuxième numéro, $Z_2 = 1$ $Y_2 = 2$. Donc $Y_2(\Omega) = \{1; 2\}$ et on note que

$$Y_2 = Y_1 + Z_1 = 1 + Z_1.$$

Ainsi,

$$\begin{aligned} \mathbb{P}(Y_2 = 1) &= \mathbb{P}(1 + Z_1 = 1) = \mathbb{P}(Z_1 = 0) = \frac{1}{N} \\ \mathbb{P}(Y_2 = 2) &= \mathbb{P}(1 + Z_1 = 2) = \mathbb{P}(Z_1 = 1) = \frac{N - 1}{N}. \end{aligned}$$

Conclusion,

$$Y_2(\Omega) = \{1; 2\}, \quad \mathbb{P}(Y_2 = 1) = \frac{1}{N}, \quad \mathbb{P}(Y_2 = 2) = \frac{N - 1}{N}.$$

- (a) L'évènement $(Y_2 = 1, Y_3 = 1)$ correspond à n'avoir pioché que un seul numéro lors des trois premiers tirages et donc d'avoir pioché le même numéro qu'au tirage 1 et 2 i.e $Y_2 = 1$ et de repiocher le même numéro au tirage 3 : $Z_3 = 0$. Ainsi,

$$(Y_2 = 1, Y_3 = 1) = (Y_2 = 1, Z_3 = 0).$$

- (b) Par la question précédente,

$$\mathbb{P}(Y_2 = 1, Y_3 = 1) = \mathbb{P}(Y_2 = 1, Z_3 = 0) = \mathbb{P}(Z_3 = 0 \mid Y_2 = 1) \mathbb{P}(Y_2 = 1).$$



On sait que $\mathbb{P}(Y_2 = 1) = \frac{1}{N}$. De plus si $(Y_2 = 1)$, alors en deux tirages une seule boule a été tirée et il reste $N - 1$ boules non tirées. Donc

$$\mathbb{P}(Z_3 = 0 \mid Y_2 = 1) = \frac{1}{N}.$$

Conclusion,

$$\mathbb{P}(Y_2 = 1, Y_3 = 1) = \frac{1}{N^2}$$

4. On sait que $Y_2(\Omega) = \llbracket 1; 2 \rrbracket$. En trois pioches, nous avons pu obtenir 1, 2 ou 3 numéros distincts. Donc $Y_3(\Omega) = \llbracket 1; 3 \rrbracket$. De même que dans les questions précédentes, on a

$$\mathbb{P}(Y_2 = 1, Y_3 = 2) = \mathbb{P}(Y_2 = 1, Z_3 = 1) = \mathbb{P}(Z_3 = 1 \mid Y_2 = 1) \mathbb{P}(Y_2 = 1) = \frac{N-1}{N} \frac{1}{N} = \frac{N-1}{N^2}.$$

$(Y_2 = 1, Y_3 = 3)$ signifie avoir eu deux fois le même numéro lors de deux premiers tirages et d'obtenir 3 numéros distincts au bout du troisième tirage. Il faudrait avoir obtenu deux numéros distincts lors du tirage 3 ce qui est impossible :

$$\mathbb{P}(Y_2 = 1, Y_3 = 3) = 0.$$

En procédant de même pour les autres probabilités, on obtient :

$Y_2 \backslash Y_3$	1	2	3
1	$\frac{1}{N^2}$	$\frac{N-1}{N^2}$	0
2	0	$\frac{2(N-1)}{N^2}$	$\frac{(N-1)(N-2)}{N^2}$

NB : on vérifie bien que $\frac{1}{N^2} + \frac{N-1}{N^2} + \frac{2(N-1)}{N^2} + \frac{(N-1)(N-2)}{N^2} = \frac{1+N-1+2N-2+N^2-3N+2}{N^2} = 1$, OK!

5. Puisque $(Y_2 = i)_{i \in \llbracket 1; 2 \rrbracket}$ forme un système complet, par la formule des probabilités totales,

$$\mathbb{P}(Y_3 = 1) = \mathbb{P}(Y_3 = 1, Y_2 = 1) + \mathbb{P}(Y_3 = 1, Y_2 = 2).$$

Par la loi conjointe, on en déduit que

$$\mathbb{P}(Y_3 = 1) = \frac{1}{N^2} + 0 = \frac{1}{N^2}.$$

De même,

$$\mathbb{P}(Y_3 = 2) = \mathbb{P}(Y_3 = 2, Y_2 = 1) + \mathbb{P}(Y_3 = 2, Y_2 = 2) = \frac{N-1}{N^2} + \frac{2(N-1)}{N^2} = \frac{3(N-1)}{N^2}$$

$$\mathbb{P}(Y_3 = 3) = \mathbb{P}(Y_3 = 3, Y_2 = 1) + \mathbb{P}(Y_3 = 3, Y_2 = 2) = 0 + \frac{(N-1)(N-2)}{N^2} = \frac{(N-1)(N-2)}{N^2}.$$

Conclusion, la loi marginale de Y_3 est bien donnée par $Y_3(\Omega) = \llbracket 1; 3 \rrbracket$ et

k	1	2	3
$\mathbb{P}(Y_3 = k)$	$\frac{1}{N^2}$	$\frac{3(N-1)}{N^2}$	$\frac{(N-1)(N-2)}{N^2}$

NB : à nouveau, $\frac{1}{N^2} + \frac{3(N-1)}{N^2} + \frac{(N-1)(N-2)}{N^2} = \frac{1+3N-3+N^2-3N+2}{N^2} = 1$.



6. Par définition,

$$\begin{aligned}\mathbb{E}(Y_3) &= \sum_{k=1}^3 k\mathbb{P}(Y_3 = k) \\ &= \frac{1}{N^2} + 2\frac{3(N-1)}{N^2} + 3\frac{(N-1)(N-2)}{N^2} \\ &= \frac{1 + 6N - 6 + 3N^2 - 9N + 6}{N^2} \\ &= \frac{3N^2 - 3N + 1}{N^2}.\end{aligned}$$

Conclusion,

$$\boxed{\mathbb{E}(Y_3) = \frac{3N^2 - 3N + 1}{N^2}.$$

7. Par définition,

$$\text{Cov}(Y_2, Y_3) = \mathbb{E}(Y_2 Y_3) - \mathbb{E}(Y_2)\mathbb{E}(Y_3).$$

D'une part,

$$\mathbb{E}(Y_2 Y_3) = \sum_{(i,j) \in \llbracket 1;2 \rrbracket \times \llbracket 1;3 \rrbracket} ij\mathbb{P}(Y_2 = i, Y_3 = j).$$

Donc d'après la loi conjointe,

$$\begin{aligned}\mathbb{E}(Y_2 Y_3) &= 1 \times 1 \times \frac{1}{N^2} + 1 \times 2 \times \frac{N-1}{N^2} + 0 + 0 + 2 \times 2 \times \frac{2(N-1)}{N^2} + 2 \times 3 \times \frac{(N-1)(N-2)}{N^2} \\ &= \frac{1 + 2N - 2 + 8N - 8 + 6N^2 - 18N + 12}{N^2} \\ &= \frac{6N^2 - 8N + 3}{N^2}.\end{aligned}$$

D'autre part, par la loi de Y_2 donnée en question 2.

$$\mathbb{E}(Y_2) = \frac{1}{N} + 2 \times \frac{N-1}{N} = \frac{2N-1}{N}.$$

Finalement, d'après la question précédente,

$$\begin{aligned}\text{Cov}(Y_2, Y_3) &= \frac{6N^2 - 8N + 3}{N^2} - \frac{2N-1}{N} \frac{3N^2 - 3N + 1}{N^2} \\ &= \frac{6N^3 - 8N^2 + 3N - 6N^3 + 6N^2 - 2N + 3N^2 - 3N + 1}{N^3} \\ &= \frac{N^2 - 2N + 1}{N^3} \\ &= \frac{(N-1)^2}{N^3}.\end{aligned}$$

Conclusion,

$$\text{Cov}(Y_2, Y_3) = \frac{(N-1)^2}{N^3}.$$

8. Puisque $N \geq 2$, par la question précédente, $\text{Cov}(Y_2, Y_3) = \frac{(N-1)^2}{N^3} > 0$. Donc les variables Y_2 et Y_3 ne sont pas corrélées. En particulier (on rappelle que la réciproque est fautive),

$\boxed{\text{les variables } Y_2 \text{ et } Y_3 \text{ ne sont pas indépendantes.}}$

**Partie 2 : Loi de Z_n**

Soit $n \in \mathbb{N}^*$. On pose $m = \min(n, N)$.

9. Y_n désigne le nombre de nouveaux numéros obtenus lors de $n \geq 1$ tirages. Nécessairement, Y_n est un nombre strictement positif : $Y_n \geq 0$. De plus en n tirages on ne peut pas obtenir plus que n numéros : $Y_n \leq n$ ni plus de numéros que ne contient l'urne : $Y_n \leq N$. Donc $Y_n \leq m$. Conclusion,

$$Y_n(\Omega) \subseteq \llbracket 1; m \rrbracket.$$

10. Soit $k \in \llbracket 1; m \rrbracket$. Supposons ($Y_n = k$). Alors en n tirages nous avons obtenu exactement k numéros distincts. Il reste donc dans l'urne $N - k$ numéros non piochés et k numéros déjà piochés. Le tirage étant uniforme, la probabilité d'obtenir un nouveau numéro est alors,

$$\forall k \in \llbracket 1; m \rrbracket, \quad \mathbb{P}(Z_{n+1} = 1 \mid Y_n = k) = \frac{N - k}{N}.$$

11. La famille $(Y_n = k)_{k \in \llbracket 1; m \rrbracket}$ forme un système complet d'évènements. Donc par la formule des probabilités totales,

$$\begin{aligned} \mathbb{P}(Z_{n+1} = 1) &= \sum_{k=1}^m \mathbb{P}(Z_{n+1} = 1 \mid Y_n = k) \mathbb{P}(Y_n = k) \\ &= \sum_{k=1}^m \frac{N - k}{N} \mathbb{P}(Y_n = k) && \text{par la question précédente} \\ &= \sum_{k=1}^m \left(1 - \frac{k}{N}\right) \mathbb{P}(Y_n = k) \\ &= \sum_{k=1}^m \mathbb{P}(Y_n = k) - \frac{1}{N} \sum_{k=1}^m k \mathbb{P}(Y_n = k) \\ &= 1 - \frac{1}{N} \mathbb{E}(Y_n). \end{aligned}$$

Conclusion,

$$\mathbb{P}(Z_{n+1} = 1) = 1 - \frac{1}{N} \mathbb{E}(Y_n).$$

12. Chaque Z_i compte si le tirage i a apporté un nouveau numéro ou non tandis que Y_n compte le nombre total de nouveaux numéros. Par conséquent,

$$Y_n = \sum_{k=1}^n Z_k.$$

13. Par la question 11.

$$\begin{aligned} \mathbb{P}(Z_{n+1} = 1) &= 1 - \frac{1}{N} \mathbb{E}(Y_n) \\ &= 1 - \frac{1}{N} \mathbb{E}\left(\sum_{k=1}^n Z_k\right) && \text{par la question précédente} \\ &= 1 - \frac{1}{N} \sum_{k=1}^n \mathbb{E}(Z_k) && \text{par linéarité de l'espérance.} \end{aligned}$$



On note que pour tout $k \in \mathbb{N}^*$, Z_k n'a que deux issues 0 et 1. Donc Z_k suit une loi de Bernoulli. Notons p_k son paramètre. Dès lors, on sait que

$$\mathbb{E}(Z_k) = p_k = \mathbb{P}(Z_k = 1).$$

Conclusion,

$$\mathbb{P}(Z_{n+1} = 1) = 1 - \frac{1}{N} \sum_{k=1}^n \mathbb{P}(Z_k = 1).$$

14. Procédons par récurrence forte. Posons pour tout $n \in \mathbb{N}^*$, $\mathcal{P}(n)$: « $\mathbb{P}(Z_n = 1) = \left(1 - \frac{1}{N}\right)^{n-1}$ ».

Initialisation. Si $n = 1$, on a $Z_1 = 1$. Donc

$$\mathbb{P}(Z_n = 1) = 1.$$

D'autre part, $\left(1 - \frac{1}{N}\right)^{n-1} = \left(1 - \frac{1}{N}\right)^0 = 1$. Donc $\mathcal{P}(1)$ est vraie.

Hérédité. Soit $n \in \mathbb{N}^*$. Supposons que pour tout $k \in \llbracket 1; n \rrbracket$, $\mathcal{P}(k)$ est vraie et montrons $\mathcal{P}(n+1)$. Par la question précédente,

$$\mathbb{P}(Z_{n+1} = 1) = 1 - \frac{1}{N} \sum_{k=1}^n \mathbb{P}(Z_k = 1).$$

Par l'hypothèse de récurrence, $\forall k \in \llbracket 1; n \rrbracket$, $\mathbb{P}(Z_k = 1) = \left(1 - \frac{1}{N}\right)^{k-1}$. Ainsi,

$$\begin{aligned} \mathbb{P}(Z_{n+1} = 1) &= 1 - \frac{1}{N} \sum_{k=1}^n \left(1 - \frac{1}{N}\right)^{k-1} \\ &= 1 - \frac{1}{N} \sum_{k=1}^n \left(1 - \frac{1}{N}\right)^{k-1} \\ &= 1 - \frac{1}{N} \sum_{k=0}^{n-1} \left(1 - \frac{1}{N}\right)^k && \text{en posant } \tilde{k} = k - 1 \\ &= 1 - \frac{1}{N} \frac{1 - \left(1 - \frac{1}{N}\right)^n}{1 - \left(1 - \frac{1}{N}\right)} && \text{car on a une somme géométrique de raison } 1 - \frac{1}{N} \neq 1 \\ &= 1 - \frac{1}{N} \frac{1 - \left(1 - \frac{1}{N}\right)^n}{\frac{1}{N}} \\ &= 1 - \left[1 - \left(1 - \frac{1}{N}\right)^n\right] \\ &= \left(1 - \frac{1}{N}\right)^n. \end{aligned}$$

On obtient bien $\mathcal{P}(n+1)$.

Conclusion, pour tout $n \in \mathbb{N}^*$, $\mathcal{P}(n)$ est vraie :

$$\forall n \in \mathbb{N}^*, \quad \mathbb{P}(Z_n = 1) = \left(1 - \frac{1}{N}\right)^{n-1}.$$

**Partie 3 : Autour de Y_N**

15. Soit $n \in \mathbb{N}^*$. D'après la question 12. on a

$$\begin{aligned}
 \mathbb{E}(Y_n) &= \mathbb{E}\left(\sum_{k=1}^n Z_k\right) = \sum_{k=1}^n \mathbb{E}(Z_k) && \text{par linéarité de l'espérance} \\
 &= \sum_{k=1}^n \mathbb{P}(Z_k = 1) && \text{car } Z_k \text{ suit une loi de Bernoulli} \\
 &= \sum_{k=1}^n \left(1 - \frac{1}{N}\right)^{k-1} \\
 &= \sum_{k=0}^{n-1} \left(1 - \frac{1}{N}\right)^k \\
 &= \frac{1 - \left(1 - \frac{1}{N}\right)^n}{\frac{1}{N}} \\
 &= N \left(1 - \left(1 - \frac{1}{N}\right)^n\right).
 \end{aligned}$$

Conclusion,

$$\forall n \in \mathbb{N}^*, \quad \mathbb{E}(Y_n) = N \left(1 - \left(1 - \frac{1}{N}\right)^n\right).$$

En particulier, pour $n = 3$,

$$\begin{aligned}
 \mathbb{E}(Y_3) &= N \left(1 - \left(1 - \frac{1}{N}\right)^3\right) = N \left[1 - \left(1 - \frac{3}{N} + \frac{3}{N^2} - \frac{1}{N^3}\right)\right] \\
 &= N \frac{3N^2 - 3N + 1}{N^3} \\
 &= \frac{3N^2 - 3N + 1}{N^2}
 \end{aligned}$$

Conclusion, on retrouve bien le résultat de la question 6.

$$\mathbb{E}(Y_3) = \frac{3N^2 - 3N + 1}{N^2}.$$

16. Par la question précédente,

$$\mathbb{E}(Y_N) = N \left(1 - \left(1 - \frac{1}{N}\right)^N\right) = N \left(1 - e^{N \ln\left(1 - \frac{1}{N}\right)}\right).$$

Donc

$$\begin{aligned}
 \mathbb{E}(Y_N) &\underset{N \rightarrow +\infty}{=} N \left(1 - e^{N\left(-\frac{1}{N} + o\left(\frac{1}{N}\right)\right)}\right) \\
 &= N \left(1 - e^{-1 + o(1)}\right) \\
 &\underset{N \rightarrow +\infty}{=} N \left(1 - e^{-1} e^{o(1)}\right) \\
 &= N \left(1 - e^{-1} (1 + o(1))\right) \\
 &\underset{N \rightarrow +\infty}{=} N \left(1 - \frac{1}{e} + o(1)\right) \\
 &\underset{N \rightarrow +\infty}{=} N \left(1 - \frac{1}{e}\right) + o(N).
 \end{aligned}$$



Conclusion,

$$\mathbb{E}(Y_N) \underset{N \rightarrow +\infty}{\sim} N \left(1 - \frac{1}{e}\right).$$

On pose pour tout $n \in \mathbb{N}^*$, X_n le nombre de numéros encore jamais pioché durant les étapes 1 à n .

17. Puisque Y_N est le nombre de nouveaux numéros obtenus et X_N le nombre de numéros encore non piochés, la somme des deux donne bien le nombre de tous les numéros. On a donc

$$Y_N + X_N = N.$$

Conclusion,

$$X_N = N - Y_N.$$

18. Par la question précédente et la linéarité de l'espérance, pour tout $N \geq 2$,

$$\frac{\mathbb{E}(X_N)}{N} = \frac{\mathbb{E}(N - Y_N)}{N} = \frac{N - \mathbb{E}(Y_N)}{N} = 1 - \frac{\mathbb{E}(Y_N)}{N}.$$

Donc par la question 16.

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{E}(X_N)}{N} = 1 - \left(1 - \frac{1}{e}\right) = \frac{1}{e} < \frac{1}{2}.$$

En posant $\varepsilon = \frac{1}{2} - \frac{1}{e}$, il existe $N_0 \in \mathbb{N}^*$ tel que pour tout $N \geq N_0$,

$$\frac{1}{e} - \varepsilon \leq \frac{\mathbb{E}(X_N)}{N} \leq \frac{1}{e} + \varepsilon = \frac{1}{2}.$$

Conclusion,

$$\exists N_0 \in \mathbb{N}^*, \forall N \geq N_0, \quad \frac{\mathbb{E}(X_N)}{N} \leq \frac{1}{2}.$$

19. Soit $N \geq N_0$. On observe que

$$\mathbb{P}\left(Y_N \leq \frac{N}{3}\right) = \mathbb{P}\left(N - X_N \leq \frac{N}{3}\right) = \mathbb{P}\left(\frac{2N}{3} \leq X_N\right).$$

Par l'inégalité de Markov, car X_N est une variable aléatoire à valeurs positive,

$$\mathbb{P}\left(Y_N \leq \frac{N}{3}\right) = \mathbb{P}\left(X_N \geq \frac{2N}{3}\right) \leq \frac{3\mathbb{E}(X_N)}{2N}.$$

Donc par la question précédente, comme $N \geq N_0$,

$$\mathbb{P}\left(Y_N \leq \frac{N}{3}\right) \leq \frac{3 \cdot 1}{2 \cdot 2} = \frac{3}{4}.$$

Conclusion,

$$\forall N \geq N_0, \quad \mathbb{P}\left(Y_N \leq \frac{N}{3}\right) \leq \frac{3}{4}.$$

L'inégalité n'est pas formidable mais permet quand même de quantifier un peu que lorsque N devient assez grand, Y_N ne prend pas trop de petites valeurs.

**Partie 4 : Par les fonctions génératrices**

On fixe à nouveau $N \geq 2$. Pour tout $n \in \mathbb{N}$, on note G_n la fonction génératrice de Y_n .

20. Puisque $Y_0 = 0$. On a

$$\forall t \in \mathbb{R}, \quad G_0(t) = t^0 \mathbb{P}(Y_0 = 0) = 1.$$

Puisque $Y_1 = 1$,

$$\forall t \in \mathbb{R}, \quad G_1(t) = t^1 \mathbb{P}(Y_1 = 1) = t.$$

Enfin, par la question 2.

$$\forall t \in \mathbb{R}, \quad G_2(t) = t \mathbb{P}(Y_2 = 1) + t^2 \mathbb{P}(Y_2 = 2) = \frac{t}{N} + \frac{(N-1)t^2}{N}.$$

Conclusion,

$$\boxed{\forall t \in \mathbb{R}, \quad G_0(t) = 1, \quad G_1(t) = t, \quad G_2(t) = \frac{1}{N}t + \frac{N-1}{N}t^2 = \frac{t + (N-1)t^2}{N}.$$

21. Soit $n \in \mathbb{N}$. Soit $k \in \llbracket 0; N \rrbracket$. Supposons $1 \leq k \leq n-1$. Puisque $(Y_n = i)_{i \in \llbracket 0; m \rrbracket}$ forme un système complet d'évènements, par la formule des probabilités totales,

$$\mathbb{P}(Y_{n+1} = k) = \sum_{i=0}^n \mathbb{P}(Y_{n+1} = k \mid Y_n = i) \mathbb{P}(Y_n = i).$$

Si $(Y_n = i)$ on a obtenu i numéros distincts pendant les n premiers tirages. Dans ce cas, si l'on n'obtient pas un nouveau numéro, $Y_{n+1} = Y_n = i$ et si l'on obtient un nouveau numéro, $Y_{n+1} = Y_n + 1 = i + 1$. Donc si $k \neq i$ et $k \neq i + 1$,

$$\mathbb{P}(Y_{n+1} = k \mid Y_n = i) = 0$$

Cela correspond aussi à $i \neq k$ et $i \neq k - 1$. Alors,

$$\mathbb{P}(Y_{n+1} = k) = \mathbb{P}(Y_{n+1} = k \mid Y_n = k - 1) \mathbb{P}(Y_n = k - 1) + \mathbb{P}(Y_{n+1} = k \mid Y_n = k) \mathbb{P}(Y_n = k).$$

Si $(Y_n = k - 1)$, alors $(Y_{n+1} = k)$ correspond à avoir un nouveau numéro sachant que $k - 1$ ont déjà été obtenus :

$$\mathbb{P}(Y_{n+1} = k \mid Y_n = k - 1) = \frac{N - (k - 1)}{N} = 1 - \frac{k - 1}{N}.$$

D'autre part, si $(Y_n = k)$, alors $(Y_{n+1} = k)$ correspond à ne pas avoir un nouveau numéro sachant que k ont déjà été obtenus :

$$\mathbb{P}(Y_{n+1} = k \mid Y_n = k) = \frac{k}{N}.$$

Conclusion,

$$\boxed{\mathbb{P}(Y_{n+1} = k) = \frac{k}{N} \mathbb{P}(Y_n = k) + \left(1 - \frac{k - 1}{N}\right) \mathbb{P}(Y_n = k - 1).$$

Eventuellement quelques-uns de ces évènements sont négligeables, par exemple si $k = 0$ ou $k > n$ mais la formule reste vraie.

22. Soit $n \in \mathbb{N}$. Soit $t \in \mathbb{R}$. Par définition, puisque $Y_{n+1}(\Omega) \subseteq \llbracket 1; m \rrbracket \subseteq \llbracket 1; N \rrbracket$,

$$G_{n+1}(t) = \sum_{k=1}^N t^k \mathbb{P}(Y_{n+1} = k).$$



Là encore certains de ces évènements peuvent être négligeables. Donc par la question précédente,

$$\begin{aligned}
G_{n+1}(t) &= \sum_{k=1}^N \left[t^k \frac{k}{N} \mathbb{P}(Y_n = k) + t^k \left(1 - \frac{k-1}{N} \right) \mathbb{P}(Y_n = k-1) \right] \\
&= \frac{1}{N} \sum_{k=1}^N kt^k \mathbb{P}(Y_n = k) + \sum_{k=1}^N t^k \left(1 - \frac{k-1}{N} \right) \mathbb{P}(Y_n = k-1) \\
&= \frac{1}{N} \sum_{k=1}^N kt^k \mathbb{P}(Y_n = k) + \sum_{k=0}^{N-1} t^{k+1} \left(1 - \frac{k}{N} \right) \mathbb{P}(Y_n = k) \\
&= \frac{1}{N} \sum_{k=1}^N kt^k \mathbb{P}(Y_n = k) + \sum_{k=0}^N t^{k+1} \left(1 - \frac{k}{N} \right) \mathbb{P}(Y_n = k) \\
&= \frac{1}{N} \sum_{k=1}^N kt^k \mathbb{P}(Y_n = k) + \sum_{k=0}^N t^{k+1} \mathbb{P}(Y_n = k) - \frac{1}{N} \sum_{k=0}^N kt^{k+1} \mathbb{P}(Y_n = k) \\
&= \frac{1}{N} \sum_{k=1}^N kt^k \mathbb{P}(Y_n = k) + \sum_{k=0}^N t^{k+1} \mathbb{P}(Y_n = k) - \frac{1}{N} \sum_{k=1}^N kt^{k+1} \mathbb{P}(Y_n = k) \\
&= \frac{1-t}{N} \sum_{k=1}^N kt^k \mathbb{P}(Y_n = k) + \sum_{k=0}^N t^{k+1} \mathbb{P}(Y_n = k).
\end{aligned}$$

Conclusion,

$$G_{n+1}(t) = \frac{1-t}{N} \sum_{k=1}^N kt^k \mathbb{P}(Y_n = k) + \sum_{k=0}^N t^{k+1} \mathbb{P}(Y_n = k).$$

23. Puisque

$$G_n(t) = \sum_{k=0}^N t^k \mathbb{P}(Y_n = k).$$

Par la question précédente, on a

$$\begin{aligned}
G_{n+1}(t) &= \frac{(1-t)t}{N} \sum_{k=1}^N kt^{k-1} \mathbb{P}(Y_n = k) + t \sum_{k=0}^N t^k \mathbb{P}(Y_n = k) \\
&= \frac{(1-t)t}{N} \sum_{k=1}^N kt^{k-1} \mathbb{P}(Y_n = k) + tG_n(t).
\end{aligned}$$

D'autre part, la fonction G_n est dérivable sur \mathbb{R} en tant que fonction polynomiale et

$$G'_n(t) = \sum_{k=1}^N kt^{k-1} \mathbb{P}(Y_n = k).$$

Ainsi,

$$G_{n+1} = \frac{(1-t)t}{N} G'_n(t) + tG_n(t)$$

Conclusion,

$$G_{n+1}(t) = \frac{1}{N} t(1-t) G'_n(t) + tG_n(t).$$

24. Soit $n \in \mathbb{N}$. Par la question précédente, les fonctions G_{n+1} , G_n et G'_n étant dérivables en tant que



fonctions polynomiales,

$$\begin{aligned}\forall t \in \mathbb{R}, \quad G'_{n+1}(t) &= \frac{1}{N} (t - t^2)' G'_n(t) + \frac{1}{N} t(1-t) G''_n(t) + G_n(t) + tG'_n(t) \\ &= \frac{1}{N} (1-2t) G'_n(t) + \frac{1}{N} t(1-t) G''_n(t) + G_n(t) + tG'_n(t) \\ &= \left(\frac{1-2t}{N} + t \right) G'_n(t) + \frac{1}{N} t(1-t) G''_n(t) + G_n(t).\end{aligned}$$

En évaluant en 1,

$$G'_{n+1}(1) = \left(\frac{-1}{N} + 1 \right) G'_n(1) + 0 + G_n(1)$$

Or par le cours, on sait que $\mathbb{E}(Y_{n+1}) = G'_{n+1}(1)$, $\mathbb{E}(Y_n) = G'_n(1)$ et $G_n(1) = 1$. Finalement, on obtient bien que

$$\boxed{\forall n \in \mathbb{N}, \quad \mathbb{E}(Y_{n+1}) = \left(1 - \frac{1}{N} \right) \mathbb{E}(Y_n) + 1.}$$

25. Posons pour tout $n \in \mathbb{N}$, $a_n = \mathbb{E}(Y_n)$. Par la question précédente,

$$a_{n+1} = \left(1 - \frac{1}{N} \right) a_n + 1.$$

On reconnaît une suite arithmético-géométrique. Soit $\omega \in \mathbb{R}$.

$$\omega = \left(1 - \frac{1}{N} \right) \omega + 1 \quad \Leftrightarrow \quad \frac{1}{N} \omega = 1 \quad \Leftrightarrow \quad \omega = N.$$

Posons pour tout $n \in \mathbb{N}$, $b_n = a_n - \omega = a_n - N$. Dès lors, pour tout $n \in \mathbb{N}$,

$$\begin{aligned}b_{n+1} &= a_{n+1} - N \\ &= \left(1 - \frac{1}{N} \right) a_n + 1 - N \\ &= \left(1 - \frac{1}{N} \right) (b_n + N) + 1 - N \\ &= \left(1 - \frac{1}{N} \right) b_n + N - 1 + 1 - N \\ &= \left(1 - \frac{1}{N} \right) b_n.\end{aligned}$$

Donc $(b_n)_{n \in \mathbb{N}}$ est une suite géométrique de raison $\left(1 - \frac{1}{N} \right)$. Donc

$$\forall n \in \mathbb{N}, \quad b_n = \left(1 - \frac{1}{N} \right)^n b_0 = \left(1 - \frac{1}{N} \right)^n (a_0 - N) = \left(1 - \frac{1}{N} \right)^n (\mathbb{E}(Y_0) - N) = -N \left(1 - \frac{1}{N} \right)^n$$

D'où

$$\forall n \in \mathbb{N}, \quad a_n = b_n + N = N - N \left(1 - \frac{1}{N} \right)^n.$$

On retrouve le résultat de la question 15. (et même étendu au rang 0) :

$$\boxed{\forall n \in \mathbb{N}, \quad \mathbb{E}(Y_n) = N \left(1 - \left(1 - \frac{1}{N} \right)^n \right).}$$



Problème II - Intégration

Pour tout $x \in \mathbb{R}$, on pose

$$F(x) = \int_0^1 \frac{e^{x(1+t^2)}}{1+t^2} dt.$$

1. Soit $x \in \mathbb{R}$. La fonction $f_x : t \mapsto \frac{e^{x(1+t^2)}}{1+t^2}$ est bien définie et même continue sur \mathbb{R} car pour tout $t \in \mathbb{R}$, $1+t^2 > 0$. Donc $F(x)$ existe et ce pour tout $x \in \mathbb{R}$. Conclusion,

La fonction F est bien définie sur \mathbb{R} .

2. Soit $x \in \mathbb{R}$. Pour tout $t \in \mathbb{R}$, $\frac{e^{x(1+t^2)}}{1+t^2} \geq 0$. Donc par croissance de l'intégrale car les bornes sont dans le bon sens,

$$\forall x \in \mathbb{R}, \quad F(x) \geq 0.$$

3. Par définition,

$$F(0) = \int_0^1 \frac{1}{1+t^2} dt.$$

On reconnaît la dérivée de la fonction arctan :

$$F(0) = [\arctan(t)]_{t=0}^{t=1} = \arctan(1) - \arctan(0) = \frac{\pi}{4}.$$

Conclusion,

$$F(0) = \frac{\pi}{4}.$$

4. On observe que pour tout $t \in [0; 1]$, $1+t^2 \geq 1$ donc pour tout $x \geq 0$, $x(1+t^2) \geq x$ et

$$e^{x(1+t^2)} \geq e^x.$$

Donc puisque $1+t^2 > 0$, on obtient

$$\frac{e^{x(1+t^2)}}{1+t^2} \geq \frac{e^x}{1+t^2}.$$

Par croissance de l'intégrale, car les bornes sont dans le bon sens,

$$\forall x \in \mathbb{R}_+, \quad F(x) \geq \int_0^1 \frac{e^x}{1+t^2} dt = e^x \int_0^1 \frac{1}{1+t^2} dt = e^x \frac{\pi}{4} \quad \text{d'après la question précédente.}$$

Or $\lim_{x \rightarrow +\infty} e^x \frac{\pi}{4} = +\infty$. Conclusion, par le théorème de minoration,

$$\lim_{x \rightarrow +\infty} F(x) = +\infty.$$

5. Pour tout $t \in [0; 1]$, $1+t^2 \geq 1$ donc pour tout $x \leq 0$, $x(1+t^2) \leq x$. Par croissance de la fonction exponentielle, $0 \leq e^{x(1+t^2)} \leq e^x$. Par positivité de $1+t^2$,

$$0 \leq \frac{e^{x(1+t^2)}}{1+t^2} \leq \frac{e^x}{1+t^2}.$$

Par croissance de l'intégrale,

$$\forall x \in \mathbb{R}_-, \quad 0 \leq F(x) \leq e^x \frac{\pi}{4}.$$

Or $\lim_{x \rightarrow -\infty} e^x \frac{\pi}{4} = 0$. Donc par le théorème d'encadrement,

$$\lim_{x \rightarrow -\infty} F(x) = 0.$$



On admet le résultat suivant (cf programme de deuxième année) :

$$F \text{ est dérivable sur } \mathbb{R} \quad \text{et} \quad \forall x \in \mathbb{R}, F'(x) = \int_0^1 e^{x(1+t^2)} dt.$$

On pose

$$\forall x \in \mathbb{R}, \quad A(x) = F(-x^2) \quad \text{et} \quad B(x) = \left(\int_0^x e^{-t^2} dt \right)^2.$$

6. Puisque F est dérivable sur \mathbb{R} , par composée, A est dérivable sur \mathbb{R} et

$$\forall x \in \mathbb{R}, \quad A'(x) = -2xF'(-x^2) = -2x \int_0^1 e^{-x^2(1+t^2)} dt.$$

Conclusion, A est dérivable sur \mathbb{R} et

$$\forall x \in \mathbb{R}, \quad A'(x) = -2x \int_0^1 e^{-x^2(1+t^2)} dt.$$

7. Posons $g : t \mapsto e^{-t^2}$. La fonction g est continue sur \mathbb{R} donc par le théorème fondamental de l'analyse,

$$G : x \mapsto \int_0^x e^{-t^2} dt,$$

est bien définie et même \mathcal{C}^1 sur \mathbb{R} et est une primitive de g (et même l'unique primitive de g qui s'annule en 0). Alors, on observe que

$$\forall x \in \mathbb{R}, \quad B(x) = G(x)^2.$$

G étant dérivable sur \mathbb{R} , B l'est aussi et

$$\forall x \in \mathbb{R}, \quad B'(x) = 2G'(x)G(x) = 2g(x)G(x) = 2e^{-x^2} \int_0^x e^{-t^2} dt.$$

Conclusion, B est dérivable sur \mathbb{R} et

$$\forall x \in \mathbb{R}, \quad B'(x) = 2e^{-x^2} \int_0^x e^{-t^2} dt.$$

8. Soit $x \in \mathbb{R}^*$. Posons $u = \frac{t}{x}$ i.e. $t = ux$. La fonction $u \mapsto ux$ est \mathcal{C}^1 sur \mathbb{R} et $dt = x du$. Lorsque $t = 0$, alors $u = 0$ et lorsque $t = x$, $u = 1$. Ainsi,

$$\int_0^x e^{-t^2} dt = \int_0^1 e^{-(ux)^2} x du = x \int_0^1 e^{-u^2 x^2} du.$$

Conclusion,

$$\forall x \in \mathbb{R}^*, \quad \int_0^x e^{-t^2} dt = x \int_0^1 e^{-u^2 x^2} du.$$

9. Soit $x \in \mathbb{R}$. Par la question 7.

$$B'(x) = 2e^{-x^2} \int_0^x e^{-t^2} dt.$$

Donc par la question précédente, si $x \neq 0$,

$$B'(x) = 2x e^{-x^2} \int_0^1 e^{-u^2 x^2} du = 2x e^{-x^2} \int_0^1 e^{-t^2 x^2} dt.$$



On note que si $x = 0$, $B'(0) = 0$ et donc le résultat reste encore vrai. Ainsi,

$$\forall x \in \mathbb{R}, B'(x) = 2x e^{-x^2} \int_0^1 e^{-t^2 x^2} dt.$$

D'autre part, par la question 6.

$$\forall x \in \mathbb{R}, -A'(x) = 2x \int_0^1 e^{-x^2(1+t^2)} dt = 2x \int_0^1 e^{-x^2} e^{-x^2 t^2} dt = 2x e^{-x^2} \int_0^1 e^{-t^2 x^2} dt.$$

Conclusion,

$$\boxed{\forall x \in \mathbb{R}, \quad B'(x) = -A'(x).}$$

10. Par la question précédente, puisque \mathbb{R} est un **intervalle** :

$$\exists C \in \mathbb{R}, \forall x \in \mathbb{R}, B(x) = -A(x) + C \Leftrightarrow A(x) + B(x) = C.$$

EN particulier, en $x = 0$,

$$C = A(0) + B(0) = F(0) + 0 = \frac{\pi}{4} \quad \text{par la question 3.}$$

D'où,

$$\forall x \in \mathbb{R}, \quad A(x) + B(x) = \frac{\pi}{4}.$$

Conclusion,

$$\boxed{\forall x \in \mathbb{R}, \quad F(-x^2) + \left(\int_0^x e^{-t^2} dt \right)^2 = \frac{\pi}{4}.}$$

11. Par la question précédente,

$$\forall x \in \mathbb{R}, \quad \int_0^x e^{-t^2} dt = \pm \sqrt{\frac{\pi}{4} - F(-x^2)}.$$

Or pour tout $t \in \mathbb{R}$, $e^{-t^2} \geq 0$, donc par croissance de l'intégrale, pour tout $x \geq 0$,

$$\int_0^x e^{-t^2} dt \geq 0.$$

Ainsi,

$$\forall x \in \mathbb{R}_+, \quad \int_0^x e^{-t^2} dt = +\sqrt{\frac{\pi}{4} - F(-x^2)}$$

On sait par la question 5. que $\lim_{u \rightarrow -\infty} F(u) = 0$. Donc par composition,

$$\lim_{x \rightarrow +\infty} F(-x^2) = \lim_{u \rightarrow -\infty} F(u) = 0.$$

La fonction racine carrée étant continue en $\frac{\pi}{4}$, on en déduit que $x \mapsto \int_0^x e^{-t^2} dt$ admet une limite en $+\infty$ donnée par

$$\lim_{x \rightarrow +\infty} \int_0^x e^{-t^2} dt = \sqrt{\frac{\pi}{4}} = \frac{\sqrt{\pi}}{2}.$$

Conclusion, on obtient la demi-intégrale de Gauss

$$\boxed{\lim_{x \rightarrow +\infty} \int_0^x e^{-t^2} dt = \frac{\sqrt{\pi}}{2}.}$$

**Bonus**

A ne traiter qu'une fois le reste résolu.

On souhaite démontrer ici que F est dérivable sur \mathbb{R} et donner l'expression de sa dérivée.

12. Pour tout $t \in [0; 1]$ et tout $h \in [-1; 1]$, posons $u = h(1 + t^2)$. On observe que $1 \leq 1 + t^2 \leq 2$ donc $-2 \leq u \leq 2$. La fonction exponentielle est \mathcal{C}^2 sur \mathbb{R} donc sur $[-2; 2]$. Par l'inégalité de Taylor-Lagrange, pour tout $u \in [-2; 2]$,

$$|\exp(u) - \exp(0) - u \exp'(0)| \leq \sup_{z \in [0; u] \text{ ou } [u; 0]} |\exp''(z)| \frac{|u|^2}{2}.$$

Autrement dit,

$$|e^u - 1 - u| \leq \sup_{z \in [0; u] \text{ ou } [u; 0]} e^z \frac{u^2}{2} \leq \sup_{z \in [-2; 2]} e^z \frac{u^2}{2} = e^2 \frac{u^2}{2}.$$

En posant $u = h(1 + t^2) \in [-2; 2]$, on obtient,

$$|e^{h(1+t^2)} - 1 - h(1+t^2)| \leq e^2 \frac{h^2(1+t^2)^2}{2} \leq e^2 \frac{4h^2}{2} = 2e^2 h^2.$$

Conclusion,

$$\boxed{\forall t \in [0; 1], \forall h \in [-1; 1], \quad |e^{h(1+t^2)} - 1 - h(1+t^2)| \leq 2e^2 h^2.}$$

13. Soit $x_0 \in \mathbb{R}$. Soit $h \in [-1; 1] \setminus \{0\}$. Posons,

$$D = \frac{F(x_0 + h) - F(x_0)}{h} - \int_0^1 e^{x_0(1+t^2)} dt$$

Par définition de F ,

$$\begin{aligned} D &= \frac{1}{h} \left(\int_0^1 \frac{e^{(x_0+h)(1+t^2)}}{1+t^2} dt - \int_0^1 \frac{e^{x_0(1+t^2)}}{1+t^2} dt \right) - \int_0^1 e^{x_0(1+t^2)} dt \\ &= \int_0^1 \frac{e^{x_0(1+t^2)}}{h(1+t^2)} e^{h(1+t^2)} dt - \int_0^1 \frac{e^{x_0(1+t^2)}}{h(1+t^2)} dt - \int_0^1 e^{x_0(1+t^2)} dt \\ &= \int_0^1 \frac{e^{x_0(1+t^2)}}{h(1+t^2)} [e^{h(1+t^2)} - 1] - e^{x_0(1+t^2)} dt \\ &= \int_0^1 \frac{e^{x_0(1+t^2)}}{h(1+t^2)} [e^{h(1+t^2)} - 1 - h(1+t^2)] dt. \end{aligned}$$

Par l'inégalité triangulaire car les bornes sont dans le bon sens,

$$|D| \leq \int_0^1 \frac{e^{x_0(1+t^2)}}{h(1+t^2)} |e^{h(1+t^2)} - 1 - h(1+t^2)| dt.$$

Par la question précédente, pour tout $t \in [0; 1]$,

$$\frac{e^{x_0(1+t^2)}}{h(1+t^2)} |e^{h(1+t^2)} - 1 - h(1+t^2)| \leq \frac{e^{x_0(1+t^2)}}{h(1+t^2)} 2e^2 h^2 dt.$$

Donc par croissance de l'intégrale car les bornes sont dans le bon sens,

$$|D| \leq \int_0^1 \frac{e^{x_0(1+t^2)}}{h(1+t^2)} 2e^2 h^2 dt = 2e^2 h \underbrace{F(x_0)}_{\text{indépendant de } h}.$$



Ceci étant vrai pour tout $h \in [-1; 1] \setminus \{0\}$ et comme $\lim_{\substack{h \rightarrow 0 \\ h \neq 0}} 2e^2 h F(x_0) = 0$, on en déduit que

$$\lim_{\substack{h \rightarrow 0 \\ h \neq 0}} |D| = 0 \quad \Leftrightarrow \quad \lim_{\substack{h \rightarrow 0 \\ h \neq 0}} D = 0.$$

Conclusion,

$$\forall x_0 \in \mathbb{R}, \quad \lim_{\substack{h \rightarrow 0 \\ h \neq 0}} \frac{F(x_0 + h) - F(x_0)}{h} = \int_0^1 e^{x_0(1+t^2)} dt.$$

En en conclut que

$$\text{la fonction } F \text{ est dérivable sur } \mathbb{R} \text{ et pour tout } x \in \mathbb{R}, F'(x) = \int_0^1 e^{x(1+t^2)} dt.$$