DNS 7 : pour le vendredi 17 janvier

Le candidat encadrera ou soulignera les résultats.
N.B. : le candidat attachera la plus grande importance à la clarté, à la précision et à la concision de la rédaction. Si un candidat est amené à repérer ce qui peut lui sembler être une erreur d'énoncé, il le signalera sur sa copie et devra poursuivre sa composition en expliquant les raisons des initiatives qu'il a été amené à prendre.

Correction

Exercice 1 *(Loi de Pascal)***.**
Soit $p \in]0,1[$ et $q = 1-p$. On considère une succession d'épreuves de Bernoulli indépendantes, de même probabilité d'échec q. On définit deux suites $(S_n)_{n\geq 1}$ et $(T_n)_{n\geq 1}$ de variables aléatoires via :

— Pour tout $n \in \mathbb{N}^\star$, S_n est la variable donnant le nombre d'épreuves pour obtenir le n-ième succès.
— $T_1 = S_1$ et, pour $n \geq 2$, T_n et le nombre d'épreuves supplémentaires pour obtenir le n-ième succès après le $(n-1)$-ième succès.

1° Exprimer, pour tout $n \in \mathbb{N}^\star$, S_n en fonctions de certaines des variables T_i.

2° Pour $n \in \mathbb{N}^\star$, déterminer la loi de T_n et donner son espérance et sa variance.

3° Déterminer la loi de S_1 et la loi de S_2.

4° Pour tout $n \in \mathbb{N}^\star$, déterminer la loi de S_n.

5° En déduire, pour tout $x \in]0,1[$ et pour tout $n \in \mathbb{N}^\star$, que : $\displaystyle\sum_{k=n}^{+\infty} \binom{k-1}{n-1} x^k = \frac{x^n}{(1-x)^n}$.

Correction :

1° On a $S_n = T_1 + \ldots + T_n$

2° On a $T_1 \hookrightarrow \mathcal{G}(p)$, il en va de même pour T_2 (C'est comme si l'épreuve commençait juste après le premier succès) et pour $T_n \hookrightarrow \mathcal{G}(p)$. Ainsi $\mathbb{E}(T_n) = \frac{1}{p}$ et $\mathbb{V}(T_n) = \frac{q}{p^2}$. On peut aussi noter que les T_n sont indépendantes (car les épreuves de Bernoulli le sont et que pour chaque T_n on en prend des différentes).

3° Comme $T_1 = S_1$, on a $S_1 \hookrightarrow \mathcal{G}(p)$. On a $S_2(\Omega) \subset [\![2,+\infty[\![$, pour $k \geq 2$, comme $S_2 = T_1 + T_2$ et comme T_1 et T_2 sont indépendantes on a (d'après la FPT avec le SCE $(T_1 = i)_{i\in\mathbb{N}^\star}$) : $\mathbb{P}(S_2 = k) = \sum_{i=1}^{k-1} \mathbb{P}(T_1 = i)\mathbb{P}_{(T_1=i)}(T_1+T_2 = k) = \sum_{i=1}^{k-1} \mathbb{P}(T_1 = i)\mathbb{P}(T_2 = k-i) = \sum_{i=1}^{k-1} q^{i-1}pq^{k-i-1}p = q^{k-2}p^2 \sum_{i=1}^{k-1} 1 = (k-1)q^{k-2}p^2$.

4° On a $S_n(\Omega) \subset [\![n,+\infty[\![$. Soit $k \geq n$, l'évènement $(S_n = k)$ est l'intersection de l'évènement $E_{k-1} =$ « $(n-1)$ succès pendant les $(k-1)$ premières épreuves » et de l'évènement $A_k =$ « succès à la k-ième épreuve » qui sont indépendantes. Or $\mathbb{P}(E_{k-1}) = \binom{k-1}{n-1}p^{n-1}q^{(k-1)-(n-1)}$ et $\mathbb{P}(A_k) = p$. Ainsi $\mathbb{P}(S_n = k) = \mathbb{P}(E_{k-1})\mathbb{P}(A_k) = \binom{k-1}{n-1}p^{n-1}q^{k-n}p = \binom{k-1}{n-1}p^n q^{k-n}$.

5° On sait que $\displaystyle\sum_{k\in S_n(\Omega)} \mathbb{P}(S_n = k) = 1$, on en déduit donc, en posant $x = q$, que : $\displaystyle\sum_{k=n}^{+\infty} \binom{k-1}{n-1}(1-x)^n x^{k-n} = 1$.

Ainsi $\displaystyle\frac{(1-x)^n}{x^n} \sum_{k=n}^{+\infty} \binom{k-1}{n-1} x^k = 1$, d'où le résultat.

Exercice 2 *(*ECRICOME*)***.**
On considère un réel $p \in]0,1[$. Une entreprise dispose de N copies d'un logiciel. Une proportion p de ces CD est infectée par un virus. Il est malheureusement impossible de discerner une copie saine d'une copie contaminée.
On suppose que les nombre N peut s'écrire $N = mn$ où m et n sont des entiers > 1.
Un responsable du service statistique propose la méthode suivante pour assainir le lot.
Les N copies initiales forment la génération 0.
On prélève n CD au hasard et avec remise dans la génération 0, on les copie chacun en m exemplaires.
Les $N = mn$ CD ainsi obtenus constituent la génération 1.
On procède de même pour fabriquer la génération 2 à partir de la génération 1 etc.

Durant tout le processus, la copie d'un CD sain est saine et celle d'un CD infecté est infectée.
Le statisticien pense que si la proportion de CD infectés p est faible alors on a de bonnes chances d'obtenir un lot sain après un assez grand nombre d'opérations.
L'objet de ce problème est de confirmer ou d'infirmer cette conjecture.

Résultat préliminaire.

On considère un couple (X, Y) de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. On considère deux entiers strictement positifs u et t. On suppose que X prend ses valeurs dans $x_1, \ldots, x_u$ et Y dans $y_1, \ldots, y_t$. g désigne une fonction définie sur $\mathbb{R}$. Pour tout entier $j \in [\![1, u]\!]$, on définit le réel :

$$\mathbb{E}(g(Y)|X = x_j) = \sum_{i=1}^{t} g(y_i)\mathbb{P}(Y = y_i|X = x_j).$$

0° Montrer que $\mathbb{E}(g(Y)) = \displaystyle\sum_{j=1}^{u} \mathbb{P}(X = x_j)\mathbb{E}(g(Y)|X = x_j)$.

k désigne dans tout le problème un entier positif ou nul. On note T_k la variable aléatoire égale au nombre de CD infectés obtenus parmi les n CD tirés dans la génération k, pour constituer la génération $k + 1$.

1° a) Quelles sont les valeurs prises par la variable T_k ?
  b) Pour tout $j \in T_k(\Omega)$, déterminer la loi conditionnelle de T_{k+1} sachant $(T_k = j)$.

2° a) En déduire, à l'aide du résultat préliminaire, une relation de récurrence entre $\mathbb{E}(T_{k+1})$ et $\mathbb{E}(T_k)$.
  b) Montrer alors que : $\forall k \in \mathbb{N}$, $\mathbb{E}(T_k) = np$.

3° On considère maintenant la variable aléatoire $Z_k = T_k(n - T_k)$.
  a) En utilisant le préliminaire avec le couple (T_k, T_{k+1}) et une fonction g bien choisie, montrer que la suite de terme général $\mathbb{E}(Z_k)$ est géométrique.
  b) Donner l'expression de $\mathbb{E}(Z_k)$ en fonction de n, k et p.
  c) Montrer que $\displaystyle\lim_{k\to+\infty} \mathbb{E}(Z_k) = 0$.

4° a) Que signifie concrètement l'événement $(Z_k = 0)$?
  b) Quelle est la plus petite valeur de $j(n - j)$ quand j décrit $[\![1, n - 1]\!]$?
  c) Calculer $\displaystyle\lim_{k\to+\infty} P(Z_k = 0)$.

5° On cherche maintenant à calculer la probabilité de l'événement F : « à partir d'un certain rang, les générations ne sont constituées que de CD infectés ».
  a) Montrer que la suite $((T_k = n))_{k\geq 0}$ est une suite croissante au sens de l'inclusion.
  b) Que représente $\mathbb{P}(F)$ pour la suite $(\mathbb{P}(T_k = n))_{k\geq 0}$?

6° Soit $j \in [\![1, n - 1]\!]$. Déterminer $\displaystyle\lim_{k\to+\infty} \mathbb{P}(T_k = j)$.

7° En utilisant le résultat de la question 2°, donner la valeur de $\mathbb{P}(F)$.

8° On cherche maintenant à calculer la probabilité de l'événement G : « À partir d'un certain rang, les générations ne sont constituées que de CD sains ».
  a) Calculer $\mathbb{P}(G)$.
  b) Que pensez-vous de la méthode proposée par le statisticien ?

Correction :

0° Comme Y est finie, on a par transfert : $\mathbb{E}(g(Y)) = \displaystyle\sum_{i=1}^{t} g(y_i)\mathbb{P}(Y = y_i)$.

On remplace ensuite $\mathbb{P}(Y = y_i)$ par $\displaystyle\sum_{j=1}^{u} \mathbb{P}(X = x_j)\mathbb{P}(Y = y_i|X = x_j)$ avec la FPT appliquée avec le SCE associé à X.
On échange les sommes finies après avoir remarqué que les couples (i, j) décrivent $[\![1, t]\!] \times [\![1, u]\!]$ et le résultat suit.

1° a) Déjà $T_k(\Omega) \subset [\![0, n]\!]$. Ensuite, comme $p > 0$ il y a au moins un CD infecté que l'on peut choisir précisément 1 fois. Ainsi $(T_{k-1} = 1)$ n'est pas impossible donc $(T_k = \ell)$ se réalise par exemple lorsque l'on prend les ℓ^{res} fois l'un des m CD contaminés de la $(k-1)^{\text{e}}$ génération et des CD sains sinon (on procède avec remise, cela est possible).
Conclusion : $T_k(\Omega) = [\![0, n]\!]$.

b) Chaque choix d'un CD est une épreuve de Bernoulli de succès « le CD est infecté », donc de probabilité de succès j/n, que l'on répète n fois et de façon indépendante (avec remise) donc $T_{k+1|(T_k=j)} \hookrightarrow \mathcal{B}(n, j/n)$.

2° a) On applique le préliminaire à $X = T_k, Y = T_{k+1}$ et $g = \mathrm{Id}$:
$\mathbb{E}(T_{k+1}) = \sum_{j=0}^{n} \mathbb{P}(T_k = j)\mathbb{E}(T_{k+1}|T_k = j) = \sum_{j=0}^{n} \mathbb{P}(T_k = j)j = \mathbb{E}(T_k)$ donc $\mathbb{E}(T_{k+1}) = \mathbb{E}(T_k)$.

b) Il reste à calculer $\mathbb{E}(T_0)$. Comme $T_0 \hookrightarrow \mathcal{B}(n,p)$ *Cf.* protocole, $\mathbb{E}(T_0) = np$ donc, d'après a) : $\forall k \in \mathbb{N}$, $\mathbb{E}(T_k) = np$.

3° a) Avec $g : x \mapsto x(n-x)$ compte tenu de $\mathbb{E}(T_k) = np$, le préliminaire fournit :
$\mathbb{E}(Z_{k+1}) = \displaystyle\sum_{j=0}^{n} \mathbb{P}(T_k = j)\mathbb{E}(T_{k+1}(n - T_{k+1})|T_k = j) = n\sum_{j=0}^{n} \mathbb{P}(T_k = j)\mathbb{E}(T_{k+1})|T_k = j) - \sum_{j=0}^{n}\mathbb{P}(T_k = j)\mathbb{E}(T_{k+1}^2|T_k = j)$, en effet $\mathbb{E}(\cdot|X = x_i)$ est linéaire en sa première variable.
On a $\mathbb{E}(T_{k+1}|T_k = j) = j$ et, par la formule de König, $j\left(\frac{n-j}{n}\right) = \mathbb{V}(T_{k+1}|T_k) = \mathbb{E}(T_{k+1}^2|T_k = j) - \mathbb{E}(T_{k+1}|T_k = j)^2$ dont on tire $\mathbb{E}(T_{k+1}^2|T_k = j) = j - j^2\left(1 - \frac{1}{n}\right)$. On remplace alors pour trouver :
$\mathbb{E}(Z_{k+1}) = n\mathbb{E}(T_k) - \mathbb{E}(T_k) + \left(1 - \frac{1}{n}\right)\mathbb{E}(T_k^2) = (n-1)\mathbb{E}(T_k) - \left(1 - \frac{1}{n}\right)\mathbb{E}(T_k^2) = \frac{n-1}{n}\mathbb{E}(T_k(n - T_k)) = \frac{n-1}{n}\mathbb{E}(Z_k)$.
Conclusion : $(\mathbb{E}(Z_k))_k$ est géométrique de raison $1 - \frac{1}{n}$.

b) On a $\mathbb{E}(Z_k) = \left(1 - \frac{1}{n}\right)^k \mathbb{E}(Z_0)$. Or $\mathbb{E}(Z_0) = n\mathbb{E}(T_0) - \mathbb{E}(T_0^2) = n^2p - np(q + np) = n(n-1)pq$ où $q = 1 - p$.
Conclusion : $\forall k \in \mathbb{N}$, $\mathbb{E}(Z_k) = n(n-1)pq\left(1 - \frac{1}{n}\right)^k$.

c) conséquence immédiate de $\left|1 - \frac{1}{n}\right| < 1$.

4° a) $(Z_k = 0)$ se réalise précisément lorsque $(T_k = 0)$ ou $(T_k = n)$ est réalisé ie lorsque l'on a pioché dans la k^{e} génération de CD soit uniquement des CD sains soit uniquement des CD infectés.
Alors les générations suivantes seront elles aussi constitués de ce même unique type de CD.
Donc $(Z_k = 0)$ est l'événement « les CD de la $(k+1)^{\text{e}}$ génération sont tous sains ou tous infectés. »

b) Il s'agit d'une parabole tournée vers le bas intersectant l'axe des abscisses en 0 et n donc de maximum atteint pour $j = \frac{n+0}{2}$ et de minimum celui des valeurs en 1 et en $n-1$: $\min(1(n-1), (n-1)(n-n+1)) = n-1$.

c) Il faut (bien sûr) faire le lien avec a) et b). En particulier b) invite à calculer l'espérance de Z_k puisque par transfert on a :
$\mathbb{E}(Z_k) = \sum_{j=0}^{n} j(n-j)\mathbb{P}(T_k = j) = \sum_{j=1}^{n-1} j(n-j)\mathbb{P}(T_k = j) \geq (n-1)(1 - \mathbb{P}(T_k = 0) - \mathbb{P}(T_k = n)) = (n-1)\mathbb{P}(Z_k = 0)$.
D'où $0 \leq \mathbb{P}(Z_k = 0) \leq \mathbb{E}(Z_k)/(n-1)$ par encadrement avec 3° c) : $\displaystyle\lim_{k\to+\infty} \mathbb{P}(Z_k = 0) = 1$.

5° a) Remarque déjà faite avant.

b) Comme $F = \bigcup_{k=0}^{+\infty}(T_k = n)$, il s'agit de la limite de $\mathbb{P}(T_k = n)$ par continuité croissante de $\mathbb{P}$ compte tenu de a).

6° Pour $j \in [\![1, n-1]\!]$, on a $\mathbb{P}(T_k = j) \leq \displaystyle\sum_{j=1}^{n-1} \mathbb{P}(T_k = j) = 1 - \mathbb{P}(Z_k = 0)$, par encadrement il vient $\displaystyle\lim_{k\to+\infty} \mathbb{P}(T_k = j) = 0$.

7° On doit partir de 2° ie de $np = \mathbb{E}(T_k)$ pour obtenir une information sur $\mathbb{P}(T_k = n)$: $\mathbb{E}(T_k) = \sum_{j=0}^{n-1} j\mathbb{P}(T_k = j) + n\mathbb{P}(T_k = n)$. D'où $\mathbb{P}(T_k = n) = p$ par encadrement puisque $0 \leq \displaystyle\frac{1}{n}\sum_{j=1}^{n-1} j\mathbb{P}(T_k = j) \leq \frac{1}{n}\sum_{j=1}^{n-1} n\mathbb{P}(T_k = j) \leq 1 - \mathbb{P}(Z_k = 0)$.

8° a) On procède comme avant : $(T_k = 0)_k$ est une suite croissante d'événements d'après le protocole de construction des générations.
Ensuite $G = \bigcup_{k=0}^{+\infty}(T_k = 0)$ donc par continuité croissante de $\mathbb{P}$, $\mathbb{P}(G) = \displaystyle\lim_{k\to+\infty} \mathbb{P}(T_k = 0)$.
Un calcul analogue à celui de 7° mène à $P(G) = q$.

b) Comme $p + q = 1$, cela dépend de la petitesse de p : si p est proche de 0 alors q est proche de 1 et la méthode proposée s'avère efficace. Si au contraire p est proche de 1 ...Enfin si $p = 1/2$ ...

Exercice 3.

Soit $f : x \mapsto \displaystyle\sum_{n=1}^{+\infty} \frac{\arctan(nx)}{n^2}$.

1° Démontrer que f est définie et continue sur $\mathbb{R}$ et de classe $\mathcal{C}^1$ sur $\mathbb{R}^\star$.

2° Étudier la limite de $f'(x)$ quand x tend vers 0 en utilisant une comparaison série/intégrale.

Indication : Considérer $t \mapsto g_x(t) = \dfrac{1}{t(1 + t^2x^2)}$ pour justifier $f'(x) \underset{x\to 0}{\sim} -\ln(x)$ et conclure.

3° Interprétation graphique.

Correction :

1° On pose, pour $n \in \mathbb{N}^\star$ et $x \in \mathbb{R}$, $f_n(x) = \arctan(nx)/n^2$. la bonne définition de f revient à la convergence simple de la série de fonctions $\sum f_n$ vers f sur $\mathbb{R}$. Comme $||f_n||_\infty^{\mathbb{R}} \leq \pi/2n^2$, le critère de Riemann fournit la convergence normale de $\sum f_n$ vers f. Il s'ensuit que f est définie sur $\mathbb{R}$ et que f est continue comme limite uniforme d'une série de fonctions continues sur $\mathbb{R}$.
Pour le caractère $\mathcal{C}^1$, on étudie la série des dérivées $f_n' : x \mapsto \frac{1}{n(1+(nx)^2)}$. Cette série diverge pour $x = 0$ la convergence simple sur $\mathbb{R}^\star$ est sinon acquise par le théorème de comparaison des séries à termes positifs compte tenu de la majoration $f_n'(x) \leq \frac{1}{n^3x^2}$ quand $x \neq 0$. Il est clair que l'on a la convergence normale sur tout segment ne contenant pas 0 pour les mêmes raisons qu'avant. Le théorème de dérivation sous le signe somme s'applique : $\sum f_n$ converge vers une fonction $\mathcal{C}^1$ sur $\mathbb{R}^\star$, ie $f \in \mathcal{C}^1(\mathbb{R}^*)$.

2° On pose comme suggéré $g_x : t \mapsto 1/t(1+t^2x^2)$, ie on a remplacé n par une variable continue t. Noter que $g_x \in \mathcal{C}^0(\mathbb{R}^\star)$.

On sait aussi d'après le théorème de dérivation sous le signe somme que sur $\mathbb{R}^\star$: $f'(x) = \displaystyle\sum_{n=1}^{+\infty} f_n'(x) = \sum_{n=1}^{+\infty} g_x(n)$.

Comme g_x est décroissante sur $\mathbb{R}_+$ (inutile de dériver : $t \mapsto t$ et $t \mapsto 1+t^2x^2$ sont croissantes à **valeurs positives** donc leur produit aussi et donc son inverse ...). En particulier sur $[k, k+1]$ ($k \in \mathbb{N}$) on a $g_x(k+1) \leq g_x(t) \leq g_k(x)$ et, par croissance de l'intégrale il vient $g_x(k+1) \leq \int_k^{k+1} g_x(t)\mathrm{d}t \leq g_x(k)$ donc, en sommant pour $k \in [\![1, n]\!]$: $\displaystyle\sum_{k=1}^{n} g_x(k+1) \leq \int_1^{n+1} g_x(t)\mathrm{d}t \leq \sum_{k=1}^{n} g_x(n)$ par la relation de Chasles. Comme $0 < g_x(t) \underset{t\to+\infty}{\sim} \dfrac{1}{t^3x^2}$, g_x est intégrable sur $[1, +\infty[$ et, quand $n \to +\infty$ il vient $f'(x) - g_x(1) \leq \displaystyle\int_1^{+\infty} g_x(t)\mathrm{d}t \leq f'(x)$ ie $\displaystyle\int_1^{+\infty} g_x(t)\mathrm{d}t \leq f'(x) \leq \int_1^{+\infty} g_x(t)\mathrm{d}t + \frac{1}{1+x^2}$.

On calcule alors l'intégrale par décomposition en éléments simples : $\dfrac{1}{t(1+t^2x^2)} = \dfrac{1}{t} - \dfrac{x^2t}{1+t^2x^2}$.

D'où pour $X \geq 1$: $\displaystyle\int_1^X g_x(t)\mathrm{d}t = \left[\ln(t) - \frac{1}{2}\ln\left(1+t^2x^2\right)\right]_1^X \underset{X\to+\infty}{\longrightarrow} -\ln(x) + \frac{1}{2}\ln\left(1+x^2\right) = \int_1^{+\infty} g_x(t)\mathrm{d}t$.

Comme $\dfrac{1}{1+x^2} = \underset{x\to0^+}{o}(\ln(x))$ et $\frac{1}{2}\ln(1+x^2) = \underset{x\to0^+}{o}(\ln(x))$ (simplement car $\lim\limits_{0^+} \ln = -\infty$), il suit de l'encadrement précédent : $f'(x) \underset{x\to0^+}{\sim} -\ln(x)$.

Conclusion : $\lim\limits_{x\to0^+} f'(x) = +\infty$ et f n'est pas dérivable en 0 ; f impaire donc $\lim\limits_{x\to0^-} f'(x) = -\infty$

3° Interprétation graphique : $\mathcal{C}_f$ présente en O une tangente verticale.