## 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.

**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 : $\displaystyle\sum_{k=n}^{+\infty} \binom{k-1}{n-1} x^k = \frac{x^n}{(1-x)^n}$.

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

$k$ 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 $\lim\limits_{k\to+\infty} P(Z_k = 0)$.

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 $\lim\limits_{k\to+\infty} \mathbb{P}(T_k = j)$.

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 ?

**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.