CC2

Durée 1h00. Les documents, la calculatrice, les téléphones portables, tablettes, ordinateurs ne sont pas autorisés. La qualité de la rédaction sera prise en compte. Le barème est donné à titre indicatif.

Exercice 1. (3 points) Soit X une variable aléatoire de loi géométrique de paramètre $p \in]0,1[$.

1. Donner la loi de X.
2. Calculer la fonction caractéristique de X.

Exercice 2. (3 points) Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes toutes de loi uniforme sur l'intervalle $[0,1]$. Soit $f:[0,1]\to\mathbb{R}$ une fonction continue. Que peut-on dire de

$$\frac{f(U_1)+\ldots+f(U_n)}{n}$$

lorsque n tend vers $+\infty$?

La fonction f, continue sur le segment $[0,1]$, y est bornée. Ainsi, les variables aléatoires $(f(U_n))_{n\geq 1}$ sont indépendantes, de même loi et admettent un moment d'ordre 1. On peut leur appliquer la loi forte des grands nombres (puisque la fonction f est bornée, les variables admettent un moment d'ordre 2 et on est même dans le cas dont on a étudié la démonstration en cours) pour trouver que $\frac{f(U_1)+\ldots+f(U_n)}{n}$ converge presque sûrement, lorsque n tend vers l'infini, vers

$$\mathbb{E}[f(U_1)] = \int_0^1 f(t)dt.$$

Cette méthode est parfois utilisée pour calculer des valeurs approchées d'intégrales et s'appelle la méthode de Monte-Carlo.

Exercice 3. (6 points) On considère une suite de jets indépendants d'un dé équilibré. On désigne par X_k le résultat du k-ème jet et par Y_n le plus grand résultat observé au cours des n premiers jets :

$$Y_n = \max_{1\leq k\leq n} X_k$$

1. Montrer que $(Y_n)_n$ converge presque surement vers une constance c à préciser.

   Il est assez évident que la suite va converger presque sûrement vers 6 : Y_n vaut 6 à partir de la première observation d'un 6, et la probabilité qu'on ne fasse aucun 6 est nulle. Formalisons ceci

   $$\mathbb{P}(\exists k\in\mathbb{N} : X_k = 6) = 1$$

   Or $(\omega : \exists k\in\mathbb{N} : X_k(\omega)=6) \subset (\omega : \lim_{n\to\infty} Y_n(\omega) = 6)$, d'où

   $$\mathbb{P}\left(\lim_{n\to\infty} Y_n = 6\right) = 1$$

   et la convergence presque sûre de Y_n vers 6 .

2. On pose : $N_n = \mathrm{Card}\,\{k\leq n : X_k = 6\}$ pour tout $n\in\mathbb{N}^*$.

(a) Écrire N_n comme une somme de variables aléatoires indépendantes et identiquement distribuées et donner sa loi.

N_n est la variable aléatoire qui compte le nombre de 6 peut se réécrire $N_n = \sum_{k=1}^n \mathbb{1}_{(X_k=6)}$. La suite de variables aléatoires $\left(Z_k := \mathbb{1}_{(X_k=6)}\right)$ est indépendante et identiquement distribuée. Et $\mathbb{1}_{(X_1=6)}$ suit une loi de Bernoulli de paramètre $\mathbb{P}\left(X_1 = 6\right) = \frac{1}{6}$, elle possède donc une espérance. Donc N_n suit une loi binomiale de paramètres n et $p = 1/6$.

(b) Établir alors la convergence presque sûre de la suite $\left(\frac{N_n}{n}\right)_{n\in\mathbb{N}^*}$.

D'après la loi forte des grands nombres, on a

$$\frac{N_n}{n} = \frac{Z_1 + \ldots + Z_n}{n} \overset{p.s.}{\to} \frac{1}{6}$$

Exercice 4. (8 points) À l'approche des élections, un institut de sondage contacte successivement des individus. Notre modèle est le suivant : les appels sont indépendants et chaque individu répond qu'il va voter pour le candidat A avec probabilité p_A (et pour le candidat B avec probabilité $p_B = 1 - p_A$). Le but est d'estimer le paramètre p_A du modèle. Soit $N_A(n)$ le nombre de réponses en faveur du candidat A collectées en n appels.

1. En utilisant l'inégalité de Tchebychev, montrer la convergence en probabilité de la variable aléatoire $\frac{N_A(n)}{n}$ vers une constante à déterminer.

Introduisons une suite X_k de variables aléatoires de Bernoulli indépendantes de paramètre p_A telle que X_k prend la valeur 1 si le k-ème individu appelé se prononce pour le candidat A, et 0 dans le cas contraire. On note que $N_A(n) = \sum_{k=1}^n X_k$ et on a

$$\mathbb{E}\left[\frac{N_A(n)}{n}\right] = p_A$$
$$\mathrm{Var}\left[\frac{N_A(n)}{n}\right] = \frac{1}{n^2} n \,\mathrm{Var}\left(X_1\right) = \frac{p_A\left(1 - p_A\right)}{n}$$

donc l'inégalité de Tchebychev appliquée à la variable aléatoire $\frac{N_A(n)}{n}$ donne, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{N_A(n)}{n} - p_A\right| \geq \varepsilon\right) \leq \frac{p_A\left(1 - p_A\right)}{n\varepsilon^2}$$

Si on avait $\sum_{n=0}^{\infty} \mathbb{P}\left(\left|\frac{N_A(n)}{n} - p_A\right| \geq \varepsilon\right) < \infty$ par application du lemme de Borel-Cantelli (cf exercice 5 TD 8), on obtiendrait la convergence presque sûre de $\frac{N_A(n)}{n}$ vers p_A. Mais l'inégalité obtenue par l'inégalité obtenue ne nous donne pas une majoration par le terme général d'une série convergente, donc n'est pas suffisante pour obtenir la convergence presque sûre...

2. Montrer que $\frac{N_A(n)}{n}$ converge en fait presque sûrement vers cette constante.

D'après la loi forte des grands nombres,

$$\frac{N_A(n)}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n} \overset{p.s}{\to} p_A$$

car la suite (X_k) est iid et que X_1 admet pour espérance p_A.

3. En reprenant l'inégalité de Tchebychev, déterminer un intervalle de confiance pour le paramètre p_A.

   Utiliser l'inégalité de Tchebychev va toutefois nous permettre de construire un intervalle de confiance de niveau $1-\delta$ pour p_A, c'est-à-dire une région qui contienne p_A avec probabilité au moins $1-\delta$. Pour cela on choisit dans l'inégalité de Tchebychev ci-dessus ε tel que $\frac{p_A(1-p_A)}{n\varepsilon^2}=\delta$. Pour cette valeur de ε on a alors

$$\mathbb{P}\left(\left|\frac{N_A(n)}{n}-p_A\right|\leq\varepsilon\right)\geq 1-\delta$$

   En remarquant que $\varepsilon=\frac{1}{\sqrt{n}}\sqrt{\frac{p_A(1-p_A)}{\delta}}$, on peut poser

$$I(n,\delta)=\left[\frac{N_A(n)}{n}-\frac{1}{\sqrt{n}}\sqrt{\frac{p_A\left(1-p_A\right)}{\delta}};\frac{N_A(n)}{n}+\frac{1}{\sqrt{n}}\sqrt{\frac{p_A\left(1-p_A\right)}{\delta}}\right]$$