DEVOIR NON SURVEILLÉ – À RENDRE LE 10/10/2016

PREMIÈRE PARTIE : PROBABILITÉS

EXERCICE 1

Soit X une variable aléatoire $(\Omega,\mathscr{F},\mathbf{P}) \longrightarrow (\mathbf{N}^*,\mathscr{P}(\mathbf{N}^*))$, où $(\Omega,\mathscr{F},\mathbf{P})$ est un espace probabilisé. On suppose que $X \sim \mathscr{G}(p)$ avec $p \in]0,1[$.

Notons $X_n = (n-X)^+ = \max(0,n-X)$, $\quad Y_n = \dfrac{n+1}{1+X_n}$.

1) Calculer $\mathbf{E}(Y_n)$ sous forme d'une somme simplifiée au maximum.
Vers quoi converge $(Y_n)_{n\in\mathbf{N}}$ presque-sûrement ?

2) Montrer que

$$\forall n \in \mathbf{N}^*, \quad \forall \omega \in \Omega, \quad Y_n(\omega) \leqslant 1 + X(\omega),$$

en déduire $\lim\limits_{n\to\infty} \mathbf{E}(Y_n)$ en utilisant le théorème de convergence dominée.

3) Utiliser le résultat de la question **2)** pour calculer $\lim\limits_{n\to\infty}\left(\dfrac{n+1}{x^{n+1}}\displaystyle\sum_{k=1}^{n}\frac{x^k}{k}\right)$, avec $x > 1$.

▶CORRECTION.

1) On applique pour tout entier $n \in \mathbf{N}$ le théorème du transfert à la fonction mesurable $x \longmapsto \dfrac{n+1}{1+\max(0,n-x)}$.
On obtient

$$\begin{aligned}
\mathbf{E}(Y_n) &= \sum_{k=1}^{\infty} \frac{n+1}{1+\max(0,n-k)}\mathbf{P}(X=k)\\
&= \sum_{k=1}^{n} \frac{n+1}{1+n-k}p(1-p)^{k-1} + (n+1)p\sum_{k=n+1}^{\infty}(1-p)^{k-1}\\
&= p(n+1)\left(\sum_{k=1}^{n}\frac{1}{1+n-k}(1-p)^{k-1} + (1-p)^n\sum_{k=n+1}^{\infty}(1-p)^{k-1-n}\right)\\
&= p(n+1)\left(\sum_{k=1}^{n}\frac{1}{1+n-k}(1-p)^{k-1} + \frac{(1-p)^n}{p}\right)
\end{aligned}$$

Fixons $\omega \in \{\omega\in\Omega : X(\omega)<\infty\}$. Alors pour n assez grand $X_n(\omega) = n - X(\omega)$. Toujours pour n assez grand, on a donc $Y_n(\omega) = \dfrac{1+n}{1+n-X(\omega)}$.
Faisant $n\to\infty$ dans l'égalité précédente, on a $Y_n \to^{p.s.}_{n\to\infty} 1$ comme $\{\omega\in\Omega : X(\omega)=\infty\} = \bigcap\limits_{n\in\mathbf{N}}\{\omega\in\Omega : X(\omega)\geqslant n\}$ est négligeable.

2) Soit $\omega \in \{\omega' : n - X(\omega') \geqslant 0\}$, alors

$$1+Y_n(\omega) = 1 + \frac{n+1}{1+n-X(\omega)} = 1 + \frac{n+1-X(\omega)+X(\omega)}{1+n-X(\omega)} = 1 + \frac{X(\omega)}{1+n-X(\omega)} \leqslant 1 + X(\omega).$$

La relation se vérifie aussi sur les $\omega \in \{\omega' : n - X(\omega') < 0\}$:

$$1+Y_n(\omega) = 1 + \frac{n+1}{1+0} \leqslant 1 + X(\omega).$$

D'après ce que l'on vient de faire, et comme $1+X \in L^1(\Omega)$ (c'est une géométrique), l'hypothèse de domination est vérifiée.
Or $Y_n \to^{p.s.}_{n\to\infty} 1$, donc $\mathbf{E}(Y_n) \to_{n\to\infty} 1$, c'est terminé.

3) Reprenons l'expression de l'espérance obtenue en **(1)**.
Alors par un changement de variable,

$$\mathbf{E}(Y_n) = p(n+1)(1-p)^n\left(\sum_{k=1}^{n}\frac{1}{k}(1-p)^{-k} + \frac{1}{p}\right).$$

Fixons $x > 1$. On peut choisir $p \in]0,1[$ tel que $x = \dfrac{1}{1-p}$. Alors en appliquant le résultat de la question précédente, on a

$$\mathbf{E}(Y_n) = \frac{p(n+1)}{x^n}\left(\sum_{k=1}^{n}\frac{x^k}{k} + \frac{1}{p}\right),$$

donc

$$\lim_{n\to\infty}\left(\frac{n+1}{x^{n+1}}\sum_{k=1}^{n}\frac{x^k}{k}\right) = \lim_{n\to\infty}\left(\frac{\mathbf{E}(Y_n)}{p} - \frac{(n+1)}{px^n}\right) = \frac{1}{px} = \frac{1}{x-1}, \text{ par croissance comparée.}$$

EXERCICE 2 (L'EXPONENTIELLE EST LA SEULE LOI À ABSENCE DE MÉMOIRE)

Soit Y une variable aléatoire dans $\mathbf{R}$, telle que $Y > 0 \quad$ ps, et pour tous $s, t > 0$,

$$\mathbf{P}(Y > s + t \mid Y > s) = \mathbf{P}(Y > t).$$

On note F sa fonction de répartition.

1) Quelle est la valeur de $\mathbf{P}(Y > 0)$?
Montrer que pour tous $t \geqslant 0$ et $n \in \mathbf{N}^*$, $\mathbf{P}(Y > nt) = \mathbf{P}(Y > t)^n$ pour tous $t \geqslant 0$. En considérant $F(n)$, vérifier que $\mathbf{P}(Y > 1) \neq 1$.

2) En considérant $F(t/n)$, montrer que $\mathbf{P}(Y > t) > 0$ pour tous $t \geqslant 0$.

3) Montrer que $\mathbf{P}(Y > rt) = \mathbf{P}(Y > t)^r$ pour tous $t \geqslant 0$ et $r \geqslant 0$. ♣ INDICATION – Dans un premier temps, commencer avec les $r \in \mathbf{Q}$.

4) Déterminer F en fonction de $\mathbf{P}(Y > 1)$ et reconnaitre la loi suivie par Y.

►CORRECTION.

1) Par hypothèse $Y > 0 \quad ps$, donc $PP(Y > 0) = 1$. On établit ensuite par récurrence que pour tout entier $n \in \mathbf{N}^*$, $\mathbf{P}(Y > nt) = \mathbf{P}(Y > t)^n$.
En effet, pour $n = 1$ c'est évident.
Si la propriété est vraie au rang n, alors

$$\mathbf{P}(Y > (n+1)t) = \mathbf{P}(Y > (n+1)t \mid Y > nt)\mathbf{P}(Y > nt) + \mathbf{P}(Y > (n+1)t \mid Y \leqslant nt)\mathbf{P}(Y \leqslant nt) = \mathbf{P}(Y > (n+1)t)\mathbf{P}(Y > nt),$$

d'après l'hypothèse d'absence de mémoire. En appliquant l'hypothèse de récurrence la propriété est démontrée.
Regardons $F(n) = \mathbf{P}(Y \leqslant n)$. On sait que, comme F est une fonction de répartition, $F(n) \longrightarrow_{n\to\infty} 1$. Or d'après la question précédente $1 - F(n) = (1 - F(1))^n$. Donc si $F(1) = 0$, on a $F(n) = 0$ pour tout entier $n \geqslant 1$, ce qui est clairement absurde.
Donc $F(1) \neq 0$ et donc de manière équivalente $\mathbf{P}(Y > 1) \neq 1$.

2) Regardons $F(t/n)$. D'après **1)**, $1 - F(t) = (1 - F(t/n))^n$. Supposons que $F(t) = 1$ pour un certain $t > 0$ (ou de manière équivalente $\mathbf{P}(Y > t) = 0$ pour un certain $t \geqslant 0$).
Alors $(1 - F(t/n))^n = 1$ pour ce même t et ce pour tout entier $n \in \mathbf{N}$. Alors en faisant $n \to \infty$, on constate que nécessairement $F(0) = 1$ par continuité à droite de F et propriété sur les limites de suites géométriques, donc $Y \leqslant 0 \quad ps$, ce qui est clairement absurde.

3) On étend ensuite ce qui a été fait pour des rationnels $r = \dfrac{p}{q} \geqslant 0$. Supposons sans restriction que $p, q \geqslant 0$.
Alors $\mathbf{P}(Y > rt) = \mathbf{P}\left(Y > \dfrac{p}{q}t\right) = \mathbf{P}\left(Y > \dfrac{1}{q}t\right)^p$. Il suffit donc de démontrer que pour tout $t \geqslant 0$ et $q > 0$, on a

$$\mathbf{P}\left(Y > \frac{1}{q}t\right) = \mathbf{P}(Y > t)^{\frac{1}{q}}.$$

Or, $\mathbf{P}\left(Y > \dfrac{1}{q}t\right)^q = \mathbf{P}(Y > t)$, ce qui achève la démonstration. Or, pour tout $r \geqslant 0$, il existe une suite $(r_n)_{n\in\mathbf{N}}$ de rationnels convergeant vers r qui est positive pour n assez grand.
Une petite subtilité ici néanmoins— on a besoin de $r_n \nearrow_{n\to\infty} r$ ou $r_n \searrow_{n\to\infty} r$ pour appliquer les propriétés de limite monotone relatives aux probabilités. Quitte à extraire une sous-suite on se ramène à l'un des deux. Par exemple, supposons $r_n \nearrow_{n\to\infty} r$.
Alors, comme on a pour tout entier $n \in \mathbf{N}$

$$\{Y > r_{n+1}t\} \subset \{Y > r_n t\},$$

et

$$\bigcap_{n\in\mathbf{N}} \{Y > r_n t\} = \{Y > rt\},$$

le théorème de convergence monotone pour les probabilités donne

$$\mathbf{P}(Y > rt) =^{(!)} \lim_{n\to\infty} \mathbf{P}(Y > r_n t) = \lim_{n\to\infty} \mathbf{P}(Y > t)^{r_n} = \mathbf{P}(Y > t)^r.$$

4) Finalement pour tout $t \geqslant 0$, on a

$$F(t) = 1 - \mathbf{P}(Y > t) = 1 - \mathbf{P}(Y > 1)^t = 1 - e^{-\lambda t},$$

choisissant $\lambda = -\ln \mathbf{P}(Y > 1) > 0$ comme $\mathbf{P}(Y > 1) \in]0,1[$ d'après **2)** et **3)**.

EXERCICE 3

1) Soit $(A_n)_{n\in\mathbf{N}}$ une suite d'évènements indépendants deux à deux, telle que A_1 ne soit pas négligeable, et $\sum_{n\in\mathbf{N}} \mathbf{P}(A_n)$ diverge.

Pour tout $n \in \mathbf{N}^*$, on définit $S_n = \sum_{k=1}^{n} \mathbb{1}_{A_k}$, et $S = \sum_{k=1}^{\infty} \mathbb{1}_{A_k}$.

a. Montrer que pour tout $n \in \mathbf{N}^*$, on a :

$$\left\{S \geqslant \frac{\mathbf{E}(S_n)}{2}\right\} \supset \left\{S_n \geqslant \frac{\mathbf{E}(S_n)}{2}\right\} \supset \left\{|S_n - \mathbf{E}(S_n)| \leqslant \frac{\mathbf{E}(S_n)}{2}\right\}.$$

b. Montrer que : $\forall \varepsilon \in]0,1[, \exists n_0 \in \mathbf{N}^*, \forall n \geqslant n_0, \quad \mathbf{P}\left(S > \dfrac{\mathbf{E}(S_n)}{2}\right) \geqslant 1-\varepsilon.$

♣ INDICATION – On pourra remarquer que

$$\mathbf{P}\left\{|S_n - \mathbf{E}(S_n)| \leqslant \frac{\mathbf{E}(S_n)}{2}\right\} = 1 - \mathbf{P}\left\{|S_n - \mathbf{E}(S_n)| > \frac{\mathbf{E}(S_n)}{2}\right\},$$

et appliquer une inégalité classique du cours au dernier terme. Conclure ensuite sur l'assertion demandée.

c. Calculer $\mathbf{P}\left(\limsup\limits_{n\to\infty} A_n\right)$.

2) Soit $(X_n)_{n\in\mathbf{N}}$ une suite de variables aléatoires deux à deux indépendantes, telles que pour tout $n \in \mathbf{N}^*$,
$\mathbf{P}(X_n = 0) = 1 - \dfrac{1}{n}, \quad \mathbf{P}(X_n = n) = \dfrac{1}{n}.$
Déterminer la variable aléatoire $\limsup\limits_{n\to\infty} \dfrac{X_n}{n}$ définie pour tout $\omega \in \Omega$ par $\left(\limsup\limits_{n\to\infty} \dfrac{X_n}{n}\right)(\omega) = \limsup\limits_{n\to\infty} \dfrac{X_n(\omega)}{n}$.

♣ INDICATION – On pourra considérer la suite d'évènements $A_n = \left\{\dfrac{X_n}{n} = 1 \text{ p.s.}\right\}$ et utiliser la question **1)**.

►CORRECTION.

1) Les deux inclusions proviennent de : pour tout $n \in \mathbf{N}$,

$$S_n \leqslant S$$

puisque S_n est, ω par ω, une somme partielle croissante et

$$|S_n - \mathbf{E}(S_n)| \leqslant \frac{\mathbf{E}(S_n)}{2} \iff \frac{\mathbf{E}(S_n)}{2} \leqslant S_n \leqslant \frac{3\mathbf{E}(S_n)}{2}.$$

D'après la première partie de la question, il suffit d'établir que

$$\mathbf{P}\left\{|S_n - \mathbf{E}(S_n)| > \frac{\mathbf{E}(S_n)}{2}\right\} \longrightarrow_{n\to\infty} 0.$$

Par l'inégalité de Bienaymé-Tchebychev, on a

$$\mathbf{P}\left\{|S_n - \mathbf{E}(S_n)| > \frac{\mathbf{E}(S_n)}{2}\right\} \leqslant \frac{4\mathbf{V}(S_n)}{\mathbf{E}(S_n)^2}.$$

Or, par indépendante des évènements A_i, les variables aléatoires $\mathbb{1}_{A_i}$ sont indépendantes, donc
$\mathbf{V}(S_n) = \sum\limits_{k=1}^{n} \mathbf{V}(\mathbb{1}_{A_k}) = \sum\limits_{k=1}^{n} \left(\mathbf{P}(\mathbb{1}_{A_k}) - \mathbf{P}(\mathbb{1}_{A_k})^2\right) \leqslant \sum\limits_{k=1}^{n} \mathbf{P}(\mathbb{1}_{A_k})$. Donc

$$\frac{4\mathbf{V}(S_n)}{\mathbf{E}(S_n)^2} \leqslant \frac{4\sum_{k=1}^{n}\mathbf{P}(\mathbb{1}_{A_k})}{\left(\sum_{k=1}^{n}\mathbf{P}(\mathbb{1}_{A_k})\right)^2} = \frac{4}{\sum_{k=1}^{n}\mathbf{P}(\mathbb{1}_{A_k})} \to_{n\to\infty} 0,$$

car par hypothèse le dénominateur diverge vers ∞. On conclut alors cette question :
$\omega \in \limsup\limits_{n} A_n$ revient à dire qu'une infinité des v.a. $\mathbb{1}_{A_n}(\omega)$ vaut 1. C'est-à-dire

$$\limsup_n A_n = \{S = \infty\} = \bigcap_{n\in\mathbf{N}} \{S \geqslant n\}.$$

Or, $\mathbf{E}(S_n) \longrightarrow_{n\to\infty} \infty$ par hypothèse donc on peut démontrer facilement par double inclusion que

$$\bigcap_{n\in\mathbf{N}} \{S \geqslant n\} = \bigcap_{n\in\mathbf{N}} \left\{S \geqslant \frac{\mathbf{E}(S_n)}{2}\right\},$$

Donc par convergence monotone pour les probabilités, on a

$$\mathbf{P}\left(\limsup_n A_n\right) = \lim_{n\to\infty} \mathbf{P}(S \geqslant n) = \lim_{n\to\infty} \mathbf{P}\left(S \geqslant \frac{\mathbf{E}(S_n)}{2}\right) = 1$$

d'après ce qui précède.

2) Par hypothèse, $\sum\limits_{n\in\mathbf{N}} \mathbf{P}(A_n) = \infty$, donc $\mathbf{P}(\limsup\limits_n A_n) = 1$ d'après **1)**.
Donc comme $\mathbf{P}(\limsup\limits_n A_n) = \mathbf{P}\left(\limsup\limits_n \dfrac{X_n}{n} = 1\right) = 1$, on a que $\limsup\limits_{n\to\infty} \dfrac{X_n}{n} = 1 \quad ps.$

EXERCICE 4 (GRANDE DÉVIATION POUR LA FRÉQUENCE DE PILES)

On se propose d'établir le théorème suivant :

THÉORÈME – 0.1 Soit $p \in]0,1[$, $(X_n)_{n\in\mathbf{N}}$ une suite de variables aléatoires i.i.d. de loi $\mathscr{B}(p)$.

On note $S_n = \sum_{i=1}^{n} X_i$ et pour tout $\varepsilon \in]0, 1-p[$,

$$h_+(\varepsilon) = (p+\varepsilon)\ln\left(\frac{p+\varepsilon}{p}\right) + (1-p-\varepsilon)\ln\left(\frac{1-p-\varepsilon}{1-p}\right).$$

Alors :

(i) $h_+(\varepsilon) > 0$,

(ii) pour tout $n \geqslant 1$, $\quad \mathbf{P}\left\{\frac{S_n}{n} \geqslant p+\varepsilon\right\} \leqslant e^{-nh_+(\varepsilon)}$.

(iii) L'inégalité est optimale :

$$\lim_{n\to\infty} \frac{1}{n}\ln\left(\mathbf{P}\left\{\frac{S_n}{n} \geqslant p+\varepsilon\right\}\right) = -h_+(\varepsilon).$$

L'assertion **(iii)** signifie que S_n/n satisfait un *principe de grande déviation* à la *vitesse* n et de *fonctionnelle de taux* h_+.

1) Quelle est la loi de S_n ?

2) Montrer que pour tout $t > 0$, $\quad \mathbf{P}\left\{\frac{S_n}{n} \geqslant p+\varepsilon\right\} = \mathbf{P}\left\{e^{t(S_n - np - n\varepsilon)} \geqslant 1\right\}$.

3) En utilisant l'inégalité de MARKOV, en déduire que pour tout $t > 0$ on a

$$\mathbf{P}\left(\frac{S_n}{n} \geqslant p+\varepsilon\right) \leqslant e^{-nh}, \quad \text{où} \quad h = \sup_{t>0}\left\{t(p+\varepsilon) - \ln(1-p+pe^t)\right\}.$$

4) Calculer h et conclure quant aux points **(i)** et **(ii)**.

5) On note $k_n = \lfloor n(p+\varepsilon)\rfloor + 1$. Montrer que $\mathbf{P}(S_n \geqslant n(p+\varepsilon)) \geqslant \mathbf{P}(S_n = k_n)$.

6) En utilisant la loi de S_n et la formule de STIRLING, établir que

$$\mathbf{P}(S_n = k_n) \sim_{n\to\infty} \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{k_n(n-k_n)}}\left(\frac{np}{k_n}\right)^{k_n}\left(\frac{n(1-p)}{n-k_n}\right)^{n-k_n},$$

puis un équivalent de $\left(\frac{1}{n}\ln \mathbf{P}(S_n = k_n)\right)_{n\in\mathbf{N}}$ lorsque $n \to \infty$.

7) En analysant la convergence de chacun des termes, montrer que

$$\frac{1}{n}\ln \mathbf{P}(S_n = k_n) \longrightarrow_{n\to\infty} -h_+(\varepsilon).$$

8) Conclure sur l'optimalité énoncée en **(iii)**.

►CORRECTION.

1) On a une somme de v.a. indépendantes suivant une Bernouilli, donc S_n suit une loi $\mathscr{B}(n,p)$.

2) Soit $t > 0$. Alors $\mathbf{P}\left\{\frac{S_n}{n} \geqslant p+\varepsilon\right\} = \mathbf{P}\{tS_n \geqslant t(np+n\varepsilon)\}$, mais $\{tS_n \geqslant t(np+n\varepsilon)\} = \left\{e^{tS_n} \geqslant e^{t(np+n\varepsilon)}\right\}$, une inclusion s'obtient en passant à l'exponentielle, l'autre en prenant le logarithme.

Donc $\mathbf{P}\left\{\frac{S_n}{n} \geqslant p+\varepsilon\right\} = \mathbf{P}\left\{e^{t(S_n-np-n\varepsilon)} \geqslant 1\right\}$.

3) Remarquons déjà que, comme les v.a. sont iid, $\mathbf{E}\left(e^{tS_n}\right) = \mathbf{E}\left(e^{tX_1}\right)^n = \left(1-p+pe^t\right)^n$. En appliquant l'inégalité de MARKOV dans **2)**, on a

$$\mathbf{P}\left(\frac{S_n}{n} \geqslant p+\varepsilon\right) \leqslant e^{-nt(p+\varepsilon)}\mathbf{E}\left(e^{tS_n}\right) = e^{-nt(p+\varepsilon)}\left(1-p+pe^t\right)^n = e^{-nt(p+\varepsilon)}\mathbf{E}\left(e^{tS_n}\right) = e^{-nt(p+\varepsilon)+n\ln(1-p+pe^t)},$$

ensuite on choisit comme le membre de gauche ne dépend pas de t on choisit le meilleur majorant à droite en prenant le sup en $t > 0$. D'où le résultat.

4) Une étude de fonction (en t) donne $h = h_+(\varepsilon)$ et $h_+(\varepsilon) > 0$.

5) L'inégalité provient de $k_n \geqslant n(p+\varepsilon)$.

6) ↓

7) Appliquant STIRLING à chacun des termes dans $\mathbf{P}(S_n = k_n) = \binom{n}{k_n}p^{k_n}(1-p)^{n-k_n}$, on obtient

$$\mathbf{P}(S_n = k_n) \sim_{n\to\infty} \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{k_n(n-k_n)}}\left(\frac{np}{k_n}\right)^{k_n}\left(\frac{n(1-p)}{n-k_n}\right)^{n-k_n},$$

car $n-k_n = n\left(1-\dfrac{k_n}{n}\right) \xrightarrow[n\to\infty]{} \infty$ et $k_n \xrightarrow[n\to\infty]{} \infty$ car $\dfrac{k_n}{n} \xrightarrow[n\to\infty]{} p+\varepsilon$ et $k_n \geqslant n(p+\varepsilon)$.

Ensuite attention, on ne peut passer au logarithme dans un équivalent sans précaution ! Mais comme $\mathbf{P}(S_n = k_n) \xrightarrow[n\to\infty]{} 0$ (vérifier la convergence du membre de droite en encadrant k_n), on peut prendre le logarithme.

On démontre ensuite que

$-\dfrac{1}{2n}\ln(2\pi) \xrightarrow[n\to\infty]{} 0$,

$\dfrac{1}{2n}\ln\left(\dfrac{n}{k_n(n-k_n)}\right) \xrightarrow[n\to\infty]{} 0$,

$\dfrac{k_n}{n}\ln\left(\dfrac{np}{k_n}\right) \xrightarrow[n\to\infty]{} (p+\varepsilon)\ln\left(\dfrac{p+\varepsilon}{p}\right)$, et

$\dfrac{n-k_n}{n}\ln\left(\dfrac{n(1-p)}{n-k_n}\right) \xrightarrow[n\to\infty]{} (1-p-\varepsilon)\ln\left(\dfrac{1-p-\varepsilon}{1-p}\right)$.

8) Découle de **5)** en passant à la limite.