EXAMEN – LE 14/12/2016, DURÉE : 3H

PROBABILITÉS & STATISTIQUE

On considère un espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$. Tous les objets aléatoires seront définis sur cet espace.

EXERCICE 1 (NOMBRE D'ENFANTS D'UNE POPULATION)

Dans une certaine population, la probabilité p_n qu'une famille ait $n \geqslant 0$ enfants est modélisée par une formule du type $p_n = a\dfrac{2^n}{n!}$ où $a > 0$.

1) Déterminer la valeur de a pour que la suite $(p_n)_{n\in\mathbf{N}}$ représente la fonction de masse d'une variable aléatoire discrète.

2) On choisit dans la suite une famille au hasard. Pour $n \geqslant 0$, on introduit E_n l'évènement « la famille choisie a n enfants » et G_n l'évènement « la famille choisie a n garçons ». On suppose dans toute la suite que les filles et garçons naissent de manière équiprobable.

- **a.** Quelle est la probabilité que la famille tirée ait au moins un garçon ?
- **b.** On suppose qu'une famille n'a aucun garçon. Quelle est la probabilité que la famille comporte deux enfants ?

►CORRECTION.

1) La condition de masse un s'écrit $\sum_{n=0}^{\infty} a\frac{2^n}{n!} = 1 = ae^2$, on choisit donc $a = e^{-2}$.

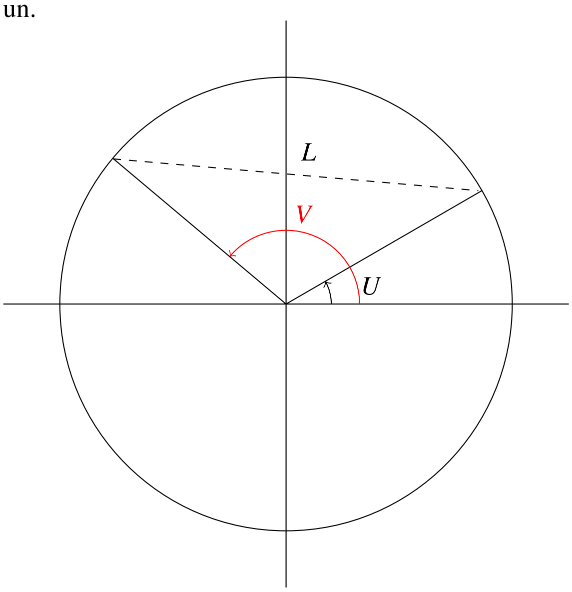
2) **a.** Notons $G = \bigcup_{n=1}^{\infty} G_n$. Calculons la probabilité de G^c.

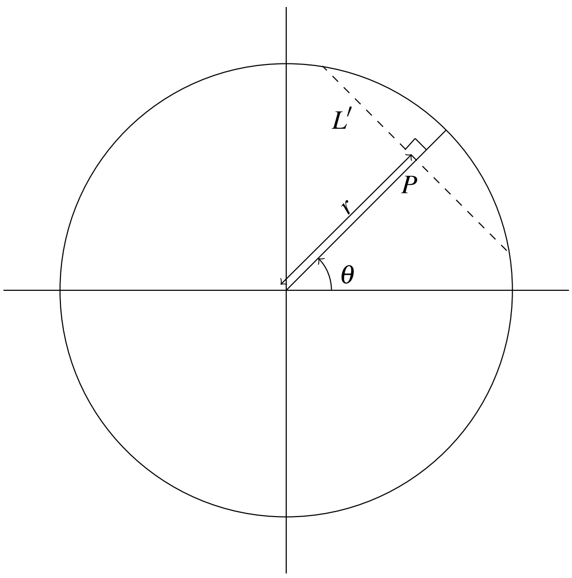
$\mathbf{P}(G^c) = \sum_{n=0}^{\infty} \mathbf{P}(G^c \mid E_n)p_n = \sum_{n=0}^{\infty} \frac{1}{2^n} p_n = e^{-1}$. Donc $\mathbf{P}(G) = 1 - e^{-1}$.

b. La probabilité cherchée est $\mathbf{P}(E_2 \mid G^c) = \dfrac{\mathbf{P}(E_2)\mathbf{P}(G^c \mid E_2)}{\mathbf{P}(G^c)} = \dfrac{1}{2e}$.

EXERCICE 2 (PARADOXE DE BERTRAND)

On étudie dans cet exercice deux façons différentes de choisir une corde reliant deux points aléatoires d'un cercle de rayon un.





1) **a.** *On considère d'abord* un mode de repérage angulaire. Soient donc U, V deux variables aléatoires suivant une uniforme $\mathscr{U}([-\pi, \pi])$, elles sont de plus supposées indépendantes. On note L la longueur de la corde reliant U et V (cf. dessin de gauche).

Montrer que L est la variable aléatoire $L = 2\sin\left|\frac{V-U}{2}\right|$. *Dans la suite, on pourra admettre cette formule même si elle n'est pas établie.*

b. On s'intéresse dans la suite à la loi de $T := \left|\frac{V-U}{2}\right|$.

Déterminer la densité du vecteur aléatoire $(V-U, V)$, en déduire celle de $V-U$.

c. Montrer que la densité de T est donnée par

$$f(t) = \frac{2}{\pi}\left(1-\frac{t}{\pi}\right)\mathbb{1}_{[0,\pi]}(t).$$

d. Calculer $\mathbf{E}(L)$, la longueur moyenne de la corde dans ce cas.

2) *On considère ensuite* le tirage suivant : on choisit d'abord une direction $\theta \in [0, 2\pi]$ suivant une $\mathscr{U}([0,2\pi])$, puis un point P au hasard sur le rayon reliant le centre du cercle au point précédemment tiré. On appelle r la distance entre le centre du cercle et P, et on suppose que $r \sim \mathscr{U}([0,1])$. Pour finir, on considère l'unique corde de milieu P, et on appelle L' sa longueur (cf. dessin de droite).

a. Démontrer que L' est la variable aléatoire $L' = 2\sqrt{1-r^2}$. *Dans la suite, on pourra admettre cette formule même si elle n'est pas établie.*

b. Calculer $\mathbf{E}(L')$, la longueur moyenne de la corde dans ce cas.

3) Dans quel cas a-t-on la plus longue corde en moyenne ?

►CORRECTION.

1) **a.** Plaçons-nous sur le sous-ensemble de Ω donné par $\{\omega \in \Omega, \quad V(\omega) \geqslant U(\omega)\}$.

Dans ce cas, en formant la hauteur (et médiane) du triangle isocèle centré en O, on a $\sin\left(\frac{V-U}{2}\right) = \frac{L}{2}$.

D'autre part, sur $\{\omega \in \Omega, \quad V(\omega) \leqslant U(\omega)\}$, on a $\sin\left(\frac{U-V}{2}\right) = \frac{L}{2}$. Ce qui établit la formule.

b. Soit $\Phi : \mathbf{R}^2 \to \mathbf{R}$ une fonction mesurable bornée. Alors

$\mathbf{E}(\Phi(V-U,V))) = \int_{[-\pi,\pi]^2} \Phi(v-u,v)\frac{1}{4\pi^2}\,\mathrm{d}u\,\mathrm{d}v$. Faisons le changement de variable donnée par $u' = v-u$ à $v \in [-\pi,\pi]$ fixé. On a alors

$\mathbf{E}(\Phi(V-U,V)) = \frac{1}{4\pi^2}\int_{[-\pi,\pi]}\left(\int_{[v-\pi,v+\pi]}\Phi(u',v)\,\mathrm{d}u'\right)\mathrm{d}v$. La densité du vecteur aléatoire $(V-U,V)$ est donc

$(u,v) \longmapsto \frac{1}{4\pi^2}\mathbb{1}_{[-\pi,\pi]}(v)\mathbb{1}_{[v-\pi,v+\pi]}(u)$. On récupère ensuite celle de $V-U$ en intégrant en la seconde variable :

$$u \longmapsto \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\mathbb{1}_{[v-\pi,v+\pi]}(u)\,\mathrm{d}v = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\mathbb{1}_{[u-\pi,u+\pi]}(v)\,\mathrm{d}v = \frac{1}{4\pi^2}\begin{cases}0 & \text{si } u \geqslant 2\pi \\ 2\pi-u & \text{si } 0 \leqslant u \leqslant 2\pi \\ 2\pi+u & \text{si } -2\pi \leqslant u \leqslant 0 \\ 0 & \text{si } u \leqslant -2\pi\end{cases}.$$

c. Soit $t \in \mathbf{R}$. Par symétrie de la loi, on a

$$\mathbf{P}\left(\left|\frac{V-U}{2}\right| \leqslant t\right) = 2\mathbf{P}(0 \leqslant V-U \leqslant 2t) = \begin{cases}\frac{2}{4\pi^2}\int_0^{2\pi}(2\pi-u)\,\mathrm{d}u = 1 & \text{si } t \geqslant \pi \\ \frac{2}{4\pi^2}\int_0^{2t}(2\pi-u)\,\mathrm{d}u = \frac{2t}{\pi^2}\left(\pi-\frac{t}{2}\right) & \text{si } 0 \leqslant t \leqslant \pi \\ 0 & \text{sinon.}\end{cases}$$

En dérivant en dehors de 0 et π la fonction de répartition ci-dessus, on trouve l'expression de la densité demandée.

d. On en déduit finalement par le théorème de transfert que

$\mathbf{E}(L) = \frac{4}{\pi}\int_0^{\pi}\sin(t)\left(1-\frac{t}{\pi}\right)\mathrm{d}t = \frac{4}{\pi}$.

2) **a.** Le théorème de Pythagore fournit $\left(\frac{L'}{2}\right)^2 + r^2 = 1$, ce qui donne ensuite la formule.

b. Par théorème de transfert, on a

$\mathbf{E}(L') = 2\int_0^1\sqrt{1-r^2}\,\mathrm{d}r$. Faisons le changement de variable $r = \cos\theta$, la fonction cos étant bijective de $[0,1]$ vers $[1, \cos 1]$, on a donc

$\mathbf{E}(L') = -2\int_{\pi/2}^{0}|\sin\theta|\sin\theta\,\mathrm{d}\theta = 2\int_0^{\pi/2}\sin^2(\theta)\,\mathrm{d}\theta = \int_0^{\pi/2}(1-\cos(2\theta))\,\mathrm{d}\theta = \frac{\pi}{2}$.

3) C'est donc dans le second cas que l'on a une longueur moyenne la plus importante.

EXERCICE 3 (VECTEURS GAUSSIENS)

Soit $(X,Y)^{\mathbf{T}}$ un vecteur gaussien de $\mathbf{R}^2$, d'espérance $\mu^{\mathbf{T}} = (\mu_1,\mu_2)^{\mathbf{T}} \in \mathbf{R}^2$ et de matrice de covariance égale à I_2 la matrice identité de taille deux, et $z = a+ib$ un nombre complexe avec $(a,b) \in \mathbf{R}^2$.

Définissons

$$X' = \text{Ré}\,(z(X+iY)), \quad Y' = \text{Im}\,(z(X+iY)), \quad A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

1) Rappeler la définition de vecteur gaussien de $\mathbf{R}^2$.

2) Exprimer $(X',Y')^{\mathbf{T}}$ en fonction de A et $(X,Y)^{\mathbf{T}}$. Montrer que $(X',Y')^{\mathbf{T}}$ est un vecteur gaussien de $\mathbf{R}^2$. Préciser l'espérance et la matrice de covariance.

3) Les variables aléatoires X' et Y' sont-elles indépendantes ?

4) **a.** Rappeler les fonctions caractéristiques de X' et Y'.

b. En déduire la fonction caractéristique de $(X',Y')^{\mathbf{T}}$.

5) À quelle condition sur μ et z, les vecteurs aléatoires $(X',Y')^{\mathbf{T}}$ et $(X,Y)^{\mathbf{T}}$ ont-ils même loi ?

►CORRECTION.

1) Voir cours

2) Après calculs on a $(X',Y')^{\mathbf{T}} = A(X,Y)^{\mathbf{T}}$, qui apparait donc comme une transformation affine du vecteur $(X,Y)^{\mathbf{T}}$ qui est donc Gaussien car $(X,Y)^{\mathbf{T}}$ l'est.
Par linéarité de l'espérance, on a $\mathbf{E}\left((X',Y')^{\mathbf{T}}\right) = A\mathbf{E}\left((X,Y)^{\mathbf{T}}\right) = A\mu$.
Pour la matrice de covariance, on a
$\mathbf{E}\left(((X',Y')^{\mathbf{T}} - A\mu)((X',Y')^{\mathbf{T}} - A\mu)^{\mathbf{T}}\right) = \mathbf{E}\left((A(X,Y)^{\mathbf{T}} - A\mu)(A(X,Y)^{\mathbf{T}} - A\mu)^{\mathbf{T}}\right) = AI_2A^{\mathbf{T}} = AA^{\mathbf{T}} = (a^2+b^2)I_2$.

3) La matrice de covariance précédente étant diagonale et le vecteur $(X',Y')^{\mathbf{T}}$ étant Gaussien, les variables aléatoires X' et Y' sont indépendantes.

4) **a.** X' et Y' sont deux variables aléatoires réelles gaussiennes de paramètres respectivement $(a\mu_1 - b\mu_2, a^2+b^2)$ et $(b\mu_1 + a\mu_2, a^2+b^2)$.

b. Par indépendance la fonction caractéristique de $(X',Y')^{\mathbf{T}}$ est définie pour tout $(x,y) \in \mathbf{R}^2$ par
$\Phi_{(X',Y')}(x,y) = \exp\left(i(a\mu_1 - b\mu_2)x - \frac{x^2}{2}(a^2+b^2)\right)\exp\left(i(b\mu_1 + a\mu_2)y - \frac{y^2}{2}(a^2+b^2)\right)$.

5) Deux vecteurs Gaussiens ont même loi si et seulement si leur espérance et matrice de covariance sont les mêmes. Ainsi on obtient comme condition : $|z|^2 = a^2+b^2 = 1, \quad A\mu = \mu$. La deuxième condition est équivalente à $a\mu_1 - b\mu_2 = \mu_1 = \mathrm{Ré}(z(\mu_1 + i\mu_2)), \quad b\mu_1 + a\mu_2 = \mu_2 = \mathrm{Im}(z(\mu_1 + i\mu_2))$

EXERCICE 4 (ÉTUDE DU MINIMUM DE PARETOS)

On considère ici une variable aléatoire réelle X qui suit une loi de Paréto de paramètres $a > 0$ et $\theta > 0$, notée $\mathscr{P}(a,\theta)$.
La loi $\mathscr{P}(a,\theta)$ admet pour densité :

$$f_{a,\theta}(x) = a\,\theta^a x^{-a-1}\,\mathbb{1}_{x>\theta}.$$

On suppose ici que le paramètre a est connu mais que θ est inconnu.
Soient $n \in \mathbf{N}^*$ et $(X_1,\ldots,X_n)$ un n-échantillon i.i.d. de X.
On pose $\widehat{\theta}_n = \min_{1\le k\le n} X_k$.

1) **a.** Calculer la fonction de répartition de $\widehat{\theta}_n$.

b. En déduire que $\widehat{\theta}_n$ suit une loi de Paréto dont on précisera les paramètres.

2) Démontrer que $\widehat{\theta}_n$ est un estimateur de θ qui converge fortement.

3) On définit pour tout entier $n \in \mathbf{N}^*$, $Z_n = n\left(\widehat{\theta}_n - \theta\right)$.

a. Montrer que la suite $(Z_n)_{n\ge1}$ converge en loi vers une variable aléatoire Z dont on déterminera la loi.

b. En déduire la valeur de : $\lim_{n\to\infty} \mathbf{P}\left[\theta\left(1+\frac{1}{n}\right) \le \widehat{\theta}_n \le \theta\left(1+\frac{2}{n}\right)\right]$.

EXERCICE 5

Pour cet exercice, on pourra utiliser sans démonstration les rappels suivants.

Rappel 1 : Une variable aléatoire réelle X suit la loi de Poisson $\mathscr{P}(\lambda)$ de paramètre $\lambda > 0$ si :

$$X \text{ est à support dans } \mathbf{N} \quad \text{et} \quad \forall k \in \mathbf{N},\ \mathbf{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Rappel 2 : Si $X \sim \mathscr{P}(\lambda)$ alors on a : $\mathbf{E}(X) = \mathbf{V}(X) = \lambda$.

Rappel 3 : $\lim\limits_{u\to 0} \dfrac{\ln(1-u)}{u} = -1.$

Soit X une variable aléatoire réelle dont la loi est inconnue et appartient au modèle $\{\mathbf{P}_\theta = \mathscr{P}(\lambda) \text{ avec } \theta = \lambda > 0\}$.
On admet que ce modèle est régulier : les hypothèses **(H1)** à **(H5)** du Cours sont vérifiées.
Etant donné $n \geq 2$ un entier fixé, on considère un n-échantillon $(X_1, \ldots, X_n)$ i.i.d. de X.

1) **a.** Démontrer que $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ est l'estimateur du maximum de vraisemblance de λ.

b. Déterminer (en détaillant vos calculs) l'information de Fisher $I_n(\lambda)$ de l'échantillon pour le paramètre λ.

2) **a.** On note $g(\lambda)$ la probabilité que X soit nulle. Donner la valeur de $g(\lambda)$ en fonction de λ.

b. On désigne par $K_n\big(g(\lambda)\big)$ la borne de Cramer-Rao du modèle pour le paramètre $g(\lambda)$.
Montrer que l'on a : $K_n\big(g(\lambda)\big) = e^{-2\lambda}\dfrac{\lambda}{n}$.

3) Soit T_n la statistique définie des deux manières équivalentes suivantes :

$$T_n := \left(1 - \frac{1}{n}\right)^{\sum_{i=1}^{n} X_i} = \exp\left[\overline{X}_n \times n \times \ln\left(1 - \frac{1}{n}\right)\right].$$

a. **(i)** Calculer $e_n := \mathbf{E}_\lambda\left[\left(1 - \dfrac{1}{n}\right)^{X_1}\right]$.
(ii) Puis montrer que T_n est un estimateur sans biais de $g(\lambda)$.

b. Montrer que l'estimateur T_n converge fortement vers $g(\lambda)$.

c. On peut démontrer (ADMIS) que T_n a les 2 propriétés suivantes :

(P1) Pour tout $\lambda > 0$, $\mathbf{V}_\lambda(T_n) = e^{-2\lambda}\left(e^{\lambda/n} - 1\right)$.

(P2) T_n est l'estimateur sans biais de $g(\lambda)$ de variance minimale i.e. que :

$$\text{pour tout estimateur } T_n' \text{ sans biais de } g(\lambda), \text{ on a : } \mathbf{V}_\lambda(T_n) \leq \mathbf{V}_\lambda(T_n'), \forall \lambda > 0.$$

En utilisant les propriétés **(P1)** et **(P2)**, répondre aux questions suivantes :

(i) L'estimateur T_n converge-t-il en moyenne quadratique vers $g(\lambda)$? Justifiez.

(ii) Pourquoi n'existe-t-il pas d'estimateur efficace de $g(\lambda)$? Justifiez.