Université de Marseille
Licence de Mathématiques, 3eme année, probabilités-Statistique
Examen du 19 mai 2016

Le polycopié du cours, les notes de cours et les notes de TD sont autorisés.

Exercice 1 (Sur le minimum de deux v.a.r. de loi exponentielle. Barème : 8 points) Soit (Ω, T, P) un espace probabilisé, X_1 une v.a.r. dont la loi est la loi exponentielle de paramètre 1 et X_2 une v.a.r. dont la loi est la loi exponentielle de paramètre 2 (soit $\lambda > 0$, on rappelle que la loi exponentielle de paramètre λ est la loi de densité f_λ avec $f_\lambda(x) = \lambda e^{-\lambda x}$ pour $x \geq 0$ et $f_\lambda(x) = 0$ pour $x < 0$). On suppose que X_1 et X_2 sont indépendantes.

Pour $\omega \in \Omega$, on pose

$$Z(\omega) = X_1(\omega) \text{ et } N(\omega) = 1, \text{ si } X_1(\omega) \leq X_2(\omega),$$
$$Z(\omega) = X_2(\omega) \text{ et } N(\omega) = 2, \text{ si } X_2(\omega) < X_1(\omega).$$

1. Montrer que, pour tout $x \in \mathbb{R}$, $\{Z \geq x\} = \{X_1 \geq x\} \cap \{X_2 \geq x\}$. Calculer la fonction de répartition de la loi de Z et en déduire la loi de Z.

Corrigé – *Soit $x \in \mathbb{R}$.*
Soit $\omega \in \Omega$. On $Z(\omega) = \min\{X_1(\omega), X_2(\omega)\}$ et donc $Z(\omega) \geq x$ si et seulement si $X_1(\omega) \geq x$ et $X_2(\omega) \geq x$. Ceci donne bien

$$\{Z \geq x\} = \{X_1 \geq x\} \cap \{X_2 \geq x\}.$$

Comme X_1 et X_2 sont indépendantes, on a donc

$$P(\{Z \geq x\}) = P(\{X_1 \geq x\})P(\{X_2 \geq x\}). \quad (1)$$

Si Y est une v.a.r. de loi exponentielle de paramètre λ ($\lambda > 0$). On a alors $P(\{Y \geq x\}) = \int_{x^+}^{\infty} \lambda e^{-\lambda t} dt = e^{-\lambda x^+}$.
La formule (1) donne donc $P(\{Z \geq x\}) = e^{-3x^+}$. La fonction de répartition de Z est donc la fonction $x \mapsto 1 - e^{-3x^+}$. Ceci prouve que Z a pour loi la loi exponentielle de paramètre 3. (On rappelle que la fonction de répartition d'une v.a.r. détermine entièrement la loi de cette v.a.r..)

2. Montrer que $P(\{X_1 = X_2\}) = 0$. En déduire que presque sûrement il existe un unique i t.q. $X_i = Z$.

Corrigé – *Comme X_1 et X_2 sont indépendantes, on $P_{(X_1,X_2)} = P_{X_1} \otimes P_{X_2}$. On a donc, avec le théorème de Fubini,*

$$P(\{X_1 = X_2\}) = \int_\Omega 1_{\{X_1=X_2\}} dP = \int_{\mathbb{R}^2} 1_{\{(x,y)\in\mathbb{R}^2, x=y\}} f_1(x)f_2(y)d(x,y) = \int_0^\infty e^{-x}(\int_x^x 2e^{-2y}dy)dx = 0.$$

En dehors de l'ensemble $\{X_1 = X_2\}$ (qui de probabilité nulle) il existe un seul i pour lequel $Z = \min\{X_1, X_2\}$. On a donc bien presque sûrement un unique i t.q. $X_i = Z$.

3. Donner la loi de N.

Corrigé – *la v.a.r. N ne prend que deux valeurs, 0 et 1. On calcule $P(\{N = 1\})$ et $P(\{N = 2\})$. On a*

$$P(\{N = 1\}) = P(\{X_1 \leq X_2\}) = \int_\Omega 1_{\{X_1\leq X_2\}} dP = \int_{\mathbb{R}^2} 1_{\{(x,y)\in\mathbb{R}^2, x\leq y\}} f_1(x)f_2(y)d(x,y)$$
$$= \int_0^\infty e^{-x}(\int_x^{+\infty} 2e^{-2y}dy)dx = \int_0^\infty e^{-3x}dx = \frac{1}{3}.$$

On a donc $P(\{N = 2\}) = \frac{2}{3}$ et la loi de N est $P_N = \frac{1}{3}\delta_1 + \frac{2}{3}\delta_2$.

4. Les v.a.r. Z et N sont elles indépendantes (justifier la réponse) ?

Corrigé – *Soit φ et ψ deux fonctions boréliennes bornée de $\mathbb{R}$ dans $\mathbb{R}$. On compare $E(\varphi(N)\psi(Z))$ et $E(\varphi(N))E(\psi(Z))$. On a*

$$E(\varphi(N)\psi(Z)) = \int_\Omega \varphi(1)\psi(X_1)1_{\{X_1\leq X_2\}}dP + \int_\Omega \varphi(2)\psi(X_2)1_{\{X_1>X_2\}}dP$$
$$= \varphi(1)\int_{\mathbb{R}^2} 1_{\{(x,y)\in\mathbb{R}^2, x\leq y\}}\psi(x)f_1(x)f_2(y)d(x,y) + \varphi(2)\int_{\mathbb{R}^2} 1_{\{(x,y)\in\mathbb{R}^2, x>y\}}\psi(y)f_1(x)f_2(y)d(x,y)$$
$$= \varphi(1)\int_0^\infty e^{-x}\psi(x)(\int_x^{+\infty} 2e^{-2y}dy)dx + \varphi(2)\int_0^\infty 2e^{-2y}\psi(y)(\int_y^{+\infty} e^{-x}dy)dx = (\varphi(1) + 2\varphi(2))\int_0^\infty e^{-3x}\psi(x)dx.$$

En prenant $\varphi = 1$ dans la formule précédente on a $E(\psi(Z)) = 3\int_0^\infty e^{-3x}\psi(x)dx$.
Comme $E(\varphi(N)) = \varphi(1)P(\{N=1\}) + \varphi(2)P(\{N=2\}) = \frac{1}{3}(\varphi(1)+2\varphi(2))$ on en déduit que $E(\varphi(N)\psi(Z)) = E(\varphi(N))E(\psi(Z))$.
Les v.a.r. Z et N sont donc indépendantes.

Exercice 2 (Somme de v.a.r.i.i.d. de loi de poisson. Barème 10 points) Soit (Ω, T, P) un espace probabilisé.

1. Soit Z une variable aléatoire réelle (v.a.r.) de loi de Poisson de paramètre m, $m > 0$ (on rappelle que la loi de Poisson de paramètre m est la loi discrète définie par $\mathbb{P}(\{Z = k\}) = \frac{m^k}{k!}e^{-m}$ pour tout $k \in \mathbb{N}$). Calculer la fonction caratétistique de Z (c'est-à-dire la fonction φ_Z, de $\mathbb{R}$ dans $\mathbb{C}$, définie par $\varphi_Z(t) = \mathbb{E}(e^{itZ})$).

   Corrigé – Soit $t \in \mathbb{R}$, on a
   $$\varphi_Z(t) = \sum_{k\in\mathbb{N}} e^{itk}P(\{Z=k\}) = \sum_{k\in\mathbb{N}} e^{itk}\frac{m^k}{k!}e^{-m} = \sum_{k\in\mathbb{N}}\frac{(me^{it})^k}{k!}e^{-m} = e^{m(e^{it}-1)}.$$

2. Soient Z_1 et Z_2 deux v.a.r. indépendantes de lois de Poisson de paramètres m_1 et m_2. Calculer la loi de $Z_1 + Z_2$.

   Corrigé – Comme Z_1 et Z_2 sont indépendantes, on a $\varphi_{Z_1+Z_2} = \varphi_{Z_1}\varphi_{Z_2}$. On a donc, pour tout $t \in \mathbb{R}$, en posant $m = m_1 + m_2$,
   $$\varphi_{Z_1+Z_2}(t) = e^{m(e^{it}-1)}.$$
   Comme la fonction caractéristique d'une v.a.r. détermine entièrement la loi de cette v.a.r., ceci montre que la loi de $Z_1 + Z_2$ est la loi de Poisson de paramètre m.

3. Soit $n \in \mathbb{N}^\star$. Montrer que la loi de Poisson de paramètre n coïncide avec la loi d'une somme de n v.a.r. indépendantes et identiquement distribuées (v.a.r.i.i.d.) dont on précisera la loi.

   Corrigé – Par récurrence sur n la question 2 montre que, pour tout $n \in \mathbb{N}^\star$, la loi d'une somme de n v.a.r.i.i.d. de loi de Poisson de paramètre 1 est la loi de Poisson de paramètre n.

Pour tout nombre réel $a \geq 0$ et tout entier n on note $\lfloor an \rfloor$ la partie entière de an et $x_n(a) = \sum_{k=0}^{\lfloor an \rfloor}\frac{n^k}{k!}$.

4. Soit $a > 0$ et $n \in \mathbb{N}^\star$.
   Montrer que $x_n(a) = e^n\mathbb{P}(\{X_n \leq an\})$ pour une variable aléatoire discrète X_n dont on précisera la loi.

   Corrigé – On prend pour X_n une v.a.r. de loi la loi de Poisson de paramètre n. On a alors
   $$e^n P(\{X_n \leq an\}) = e^n \sum_{k\leq an} P(\{X_n = k\}) = e^n \sum_{k=0}^{[an]}\frac{n^k}{k!}e^{-n} = x_n(a).$$

On se donne maintenant une suite de v.a.r.i.i.d., $(Y_i)_{n\in\mathbb{N}^\star}$, de loi de poisson de paramètre 1.

5. Soit $a \in]0,1[$. Montrer que $\lim_{n\to+\infty} e^{-n}x_n(a) = 0$.
   [Construire X_n à partir des Y_i, utiliser la question 4 et la loi faible des grands nombres.]

   Corrigé – On prend $X_n = \sum_{i=1}^n Y_i$. Comme les Y_i forment une suite de v.a.r.i.i.d et que Y_1 est de carré intégrable, la loi faible des grands nombres nous dit que X_n/n converge en probabilité vers la v.a.r. constante et égale à $E(Y_1)$. Comme $E(Y_1) = 1$, on a donc pour tout $\varepsilon > 0$
   $$\lim_{n\to+\infty} P(\{|\frac{X_n}{n} - 1| \geq \varepsilon\}) = 0.$$
   En prernant $\varepsilon = 1 - a$, on a donc $\lim_{n\to+\infty} P(\{\frac{X_n}{n} - 1 \leq a - 1\}) = 0$, ce qui donne bien
   $$\lim_{n\to+\infty} P(\{X_n \leq na\}) = 0.$$
   Pour conclure, il suffit de remarquer que $e^{-n}x_n(a) = P(\{X_n \leq na\})$.

6. Soit $a \in]1, \infty[$. Montrer que $\lim_{n\to+\infty} e^{-n} x_n(a) = 1$.

Corrigé – On reprend le raisonnement précédent avec $\varepsilon = a - 1$. On obtient $\lim_{n\to+\infty} P(\{\frac{X_n}{n} - 1 \geq a - 1\}) = 0$ et donc $\lim_{n\to+\infty} P(\{X_n \geq na\}) = 0$. On en déduit que $\lim_{n\to+\infty} P(\{X_n < na\}) = 1$ et donc aussi (par monotonie P)

$$\lim_{n\to+\infty} P(\{X_n \leq na\}) = 1.$$

On conclut encore en remarquant que $e^{-n} x_n(a) = P(\{X_n \leq na\})$.

7. Soit $a = 1$. Utiliser le théorème central limite (TCL) pour donner la valeur de $\lim_{n\to+\infty} e^{-n} x_n(a)$.

Corrigé – Le choix de X_n est le même que pour les questions précédentes. Le TCL nous donne ici que la v.a.r. $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - 1)$ converge en loi vers X où X a pour loi une loi normale centrée. Le TCL donne aussi la convergence des fonctions de répartition (car la fonction de répartition de la loi normale est continue). On a donc

$$P(\{\frac{1}{\sqrt{n}}(X_n - n)) \leq 0\}) \to P(\{X \leq 0\}) \text{ quand } n \to +\infty.$$

Comme $e^{-n} x_n(1) = P(\{X_n \leq n\}) = P(\{\frac{1}{\sqrt{n}}(X_n - n)) \leq 0\})$ et que $P(\{X \leq 0\}) = \frac{1}{2}$, on en déduit que $\lim_{n\to+\infty} e^{-n} x_n(1) = \frac{1}{2}$.

Exercice 3 (Statistique. Barème 10 points) On mesure les poids de 10 colis à envoyer, en kilogrammes. Par la suite on notera ces données comme un vecteur x donné ci-dessous (où ${}^t()$ désigne la transposée) :

$$x = {}^t (2.61, 3.24, 0.85, 4.79, 2.50, 3.42, 0.55, 1.84, 3.88, 1.26).$$

On utilise une approche de statistique inférentielle : x est une réalisation d'un vecteur aléatoire noté X. On sait d'expérience qu'un bon modèle pour ce type de données consiste à supposer que les composantes $X_1, \ldots, X_{10}$ de X sont des variables aléatoires indépendantes et identiquement distribuées (iid) de loi uniforme sur $[0, \theta]$. Ici $\theta \in \mathbb{R}_+^*$ est un paramètre déterministe inconnu dont on cherche à deviner la valeur. On fait pour cela une étude théorique à partir du vecteur aléatoire X (questions 1-2-3-4), puis on appliquera les résultats sur les données x (question 5).
On rappelle par ailleurs que si une variable U suit une loi uniforme sur $[0, \theta]$, alors $E(U) = \theta/2$, $Var(U) = \theta^2/12$ et la variable U/θ suit une loi uniforme sur $[0, 1]$.
On rappelle également la définition d'une fonction indicatrice : si A est un intervalle de $\mathbb{R}$, on note $\mathbb{I}_A$ la fonction indicatrice de A définie par $\forall y \notin A,\ \mathbb{I}_A(y) = 0$ et $\forall y \in A,\ \mathbb{I}_A(y) = 1$.

1. Trouver un estimateur de θ par la méthode des moments, en utilisant le premier moment $E(X_1)$. On notera cet estimateur $\hat{\theta}^{(1)}$.
   Déterminer le biais, la variance et le risque quadratique (ou erreur quadratique moyenne) de cet estimateur.

Corrigé – On cherche à exprimer $E(X_1)$ en fonction de θ : $E(X_1) = \theta/2$, d'après le rappel.
On inverse la relation : $\theta = 2E(X_1)$.
On définit l'estimateur en remplaçant $E(X_1)$ par $\bar{X} = \frac{\sum_{i=1}^{10} X_i}{10}$: $\hat{\theta}^{(1)} := 2\bar{X}$.
Biais : $B(\hat{\theta}^{(1)}) = E(\hat{\theta}^{(1)}) - \theta = \frac{\sum_{i=1}^{10} 2E(X_i)}{10} - \theta = \theta - \theta = 0$ par linearité de l'espérance et vu que $2E(X_i) = \theta$.
Variance : $Var(\hat{\theta}^{(1)}) = \frac{4Var(\sum_{i=1}^{10} X_i)}{100} = \frac{4\sum_{i=1}^{10} Var(X_i)}{100} = \theta^2/30$ car les variables dans la somme sont indépendantes.
Risque quadratique : $RQ(\hat{\theta}^{(1)}) = B(\hat{\theta}^{(1)})^2 + Var(\hat{\theta}^{(1)}) = \theta^2/30$ d'après la décomposition biais - variance.

2. Cette question est différente selon le site (Saint-Charles ou Luminy).

(a) (pour les étudiants de Saint-Charles) Montrer que la variable $\hat{\theta}^{(2)}$ définie par $\hat{\theta}^{(2)} := \sqrt{3\frac{\sum_{i=1}^{10} X_i^2}{10}}$ est également un estimateur de θ par la méthode des moments, mais basé cette fois ci sur le deuxième moment $E(X_1^2)$. Rappeler quelle est l'idée justifiant la construction de cet estimateur.
Montrer que le risque quadratique de $\hat{\theta}^{(2)}$ vérifie

$$RQ(\hat{\theta}^{(2)}) = C\theta^2,$$

où C ne dépend pas de θ et vaut $C = E\big(\big[\sqrt{3\frac{\sum_{i=1}^{10} Z_i^2}{10}} - 1\big]^2\big)$ avec $(Z_i)_{i=1,\dots,10}$ des variables iid de loi uniforme sur $[0,1]$.

Corrigé – On a $E(X_1^2) = Var(X_1) + E(X_1)^2 = \theta^2/12 + \theta^2/4 = \theta^2/3$, d'où l'estimateur $\hat{\theta}^{(2)}$, en procédant comme d'habitude.
Cet estimateur est fondé sur la loi des grands nombres, appliquée cette fois sur les variables X_i^2 qui sont bien iid (voir cours de proba) et de moment fini : $\frac{\sum_{i=1}^{n} X_i^2}{n}$ approxime $E(X_1^2)$ quand l'entier n est suffisamment "grand".
Risque quadratique : $RQ(\hat{\theta}^{(2)}) = E\big(\big[\sqrt{3\frac{\sum_{i=1}^{10} X_i^2}{10}} - \theta\big]^2\big) = \theta^2 E\big(\big[\sqrt{3\frac{\sum_{i=1}^{10} Z_i^2}{10}} - 1\big]^2\big)$ où $Z_i := X_i/\theta$ sont des variables iid de loi uniforme sur $[0,1]$ (voir rappel et cours de probabilités).

(b) (pour les étudiants de Luminy) Justifier en citant un résultat du cours pourquoi la variable aléatoire $\sqrt{n}\frac{\hat{\theta}^{(1)}-\theta}{2\sqrt{\hat{S}_n}}$ converge en loi vers la loi normale centrée réduite (où $\hat{S}_n$ désigne la variance empirique modifiée).
Sachant que $\mathbb{P}(Z \in [-1.96, 1.96]) = 0.95$ pour une variable normale centrée réduite Z, déduire de la question précédente un intervalle de confiance asymptotique à 95% du paramètre θ.

Corrigé – Par le théorème central limite (applicable car les X_i sont iid avec un moment d'ordre deux fini), on a $\sqrt{n}\frac{\bar{X}-E(X_1)}{\sqrt{Var(X_1)}} \longrightarrow \mathcal{N}(0,1)$ (convergence en loi).
Ainsi vu les définitions et rappels, $\sqrt{n}\frac{\hat{\theta}^{(1)}-\theta}{2\theta/\sqrt{12}} \longrightarrow \mathcal{N}(0,1)$.
De plus on sait que $\hat{S}_n$ est un estimateur consistant de $Var(X_1)$, donc $\hat{S}_n \longrightarrow \theta^2/12$ (convergence en probabilité), et ainsi $\theta/\sqrt{12\hat{S}_n} \longrightarrow 1$ (en probabilité).
Ainsi par le théorème de Slutski appliqué au produit des deux variables ci dessus, on obtient que $\sqrt{n}\frac{\hat{\theta}^{(1)}-\theta}{2\sqrt{\hat{S}_n}} \longrightarrow \mathcal{N}(0,1)$.
Enfin en raisonnant par équivalences, on voit facilement que les évènements $\sqrt{n}\frac{\hat{\theta}^{(1)}-\theta}{2\sqrt{\hat{S}_n}} \in \big[-1.96,\ 1.96\big]$ et $\theta \in \big[\hat{\theta}^{(1)} - 3.92\sqrt{\frac{\hat{S}_n}{n}}, \hat{\theta}^{(1)} + 3.92\sqrt{\frac{\hat{S}_n}{n}}\big]$ sont les mêmes. Leur probabilité tend vers 0.95 d'après l'indication, donc $\big[\hat{\theta}^{(1)} - 3.92\sqrt{\frac{\hat{S}_n}{n}}, \hat{\theta}^{(1)} + 3.92\sqrt{\frac{\hat{S}_n}{n}}\big]$ est un intervalle de confiance asymptotique à 95% du paramètre θ.

3. Montrer que la fonction de vraisemblance associée aux données x est déterminée par :

$$\forall t > 0,\ \ L(t) = \frac{1}{t^{10}} \prod_{i=1}^{10} \mathbb{I}_{[0,t]}(x_i),$$

où $\mathbb{I}$ désigne la fonction indicatrice.

Montrer que $\forall t > 0,\ \ \prod_{i=1}^{10} \mathbb{I}_{[0,t]}(x_i) = \mathbb{I}_{[\max_{i=1,\dots,10} x_i, \infty[}(t)$.

En déduire que l'estimation par maximum de vraisemblance est $\max_{i=1,\dots,10} x_i$ (et donc l'estimateur est $\hat{\theta}^{(3)} := \max_{i=1,\dots,10} X_i$).

Corrigé – *La densité de X_1 est :* $\forall y \in \mathbb{R}, \quad f_{X_1}(y) = \frac{1}{\theta}\mathbb{I}_{[0,\theta]}(y)$.
La densité de X est donc : $\forall y \in \mathbb{R}^{10}, \quad f_X(y) = \frac{1}{\theta^{10}} \prod_{i=1}^{10} \mathbb{I}_{[0,\theta]}(y_i)$.
Ainsi la fonction de vraisemblance associée aux données x est : $\forall t > 0, \quad L(t) = \frac{1}{t^{10}} \prod_{i=1}^{10} \mathbb{I}_{[0,t]}(x_i)$.

La quantité $\prod_{i=1}^{10} \mathbb{I}_{[0,t]}(x_i)$ est un produit d'indicatrices, et vaut donc 0 ou 1.
Elle vaut 1 si et seulement si $\forall i \in \{1, \dots, 10\}$, $x_i \leq t$, ie ssi $t \geq \max_{i=1,\dots,10} x_i$, d'où le résultat demandé.
Vu la forme de L, l'estimateur par maximum de vraisemblance est donc bien $\hat{\theta}^{(3)} := \max\limits_{i=1,\dots,10} X_i$.

4. On admet que $\hat{\theta}^{(3)}$ a une densité : $\forall y > 0, \quad f_{\hat{\theta}^{(3)}}(y) = \dfrac{10}{\theta^{10}} y^9 \mathbb{I}_{[0,\theta]}(y)$. Montrer alors par un calcul d'intégrale que le risque quadratique est
$$RQ(\hat{\theta}^{(3)}) = \frac{1}{66}\theta^2.$$

Corrigé – $RQ(\hat{\theta}^{(3)}) = E\left[(\hat{\theta}^{(3)} - \theta)^2\right] = \int_{\mathbb{R}} (y-\theta)^2 f_{\hat{\theta}^{(3)}}(y) dy = \int_0^\theta (y-\theta)^2 \frac{10}{\theta^{10}} y^9 dy = \theta^2 \int_0^1 10(u-1)^2 u^9 du$.

La dernière ligne s'obtient par le changement de variable $u = y/\theta$. Puis on obtient le résultat final par des calculs simples d'intégrales.

5. Application numérique : quelles sont les estimations de θ obtenues à l'aide des questions 1-2-3 ?
On admet que C défini à la question 2 vaut environ : $C = 0.021$. Ainsi, à laquelle des trois estimations donnez vous le plus de crédit, vu les expressions des risques quadratiques obtenues dans les questions 1-2-4 ? Attention les observations sont en faible nombre $n = 10$, ce qui peut donner des résultats surprenants !

Corrigé – *Les trois estimations sont respectivement environ :* 4.99, 4.88 *et* 4.79.

Les trois risques sont respectivement environ : $0.033 \times \theta^2$, $0.021 \times \theta^2$, $0.015 \times \theta^2$, *donc $\hat{\theta}^{(3)}$ est préférable car il a le plus petit risque quadratique.*

Remarques :

- *Attention de laisser θ^2 (qui est inconnu) dans les risques, et non pas ses estimations (même si normalement les valeurs sont proches). Le fait que θ soit inconnu ne pose pas de problème ici, on peut quand même comparer les risques quadratiques.*
- *Oubliez la phrase "Attention les observations sont en faible nombre $n = 10$, ce qui peut donner des résultats surprenants", l'estimateur de max de vraisemblance est en fait meilleur pour tout nombre d'observation n, et d'autant que n est grand (risque négligeable par rapport aux deux autres).*