EXAMEN PROBABILITÉS/STATISTIQUE - 2GEA

Jeudi 12 Novembre 2015

Partiel sans document (Une feuille A4 recto-verso autorisée)

Exercice 1: Probabilités (10 points)

On considère un couple de variables aléatoires (X, Y) de densité de probabilité

$$p(x,y) = \begin{cases} A \text{ si } x \geq 0 \text{ et } 0 \leq y \leq x^2 \leq 1 \\ 0 \text{ sinon} \end{cases}$$

où A est constante positive.

1. Montrer que $A = 3$.
2. Déterminer les densités marginales de X et Y. En déduire (en s'aidant lorsque cela est possible des tables de lois) les moyennes $E[X], E[Y]$ et les variances $\text{Var}[X], \text{Var}[Y]$. Les variables aléatoires X et Y sont elles indépendantes ?
3. Quelle est la loi conditionnelle de $Y|X = x$?
4. Déterminer la loi du couple (Z, T) avec $Z = \frac{Y}{X^2}$ et $T = X$. En déduire les lois marginales de Z et de T. Les variables aléatoires Z et T sont-elles indépendantes ?

Exercice 2: Estimation (6 points)

On considère une suite de variables aléatoires X_i indépendantes et de même loi géométrique à valeurs dans $\mathbb{N}^* = \{n \in \mathbb{N}, n \geq 1\}$ définie par

$$P[X = x_i] = p(1-p)^{x_i - 1}, \quad x_i = 1, 2, ...$$

On admettra qu'une telle loi est de moyenne $E[X_i] = \frac{1}{p}$ et de variance $\text{Var}(X_i) = \frac{q}{p^2}$ avec $q = 1 - p$.

1. Déterminer la vraisemblance de l'échantillon $(X_1, ..., X_n)$ et montrer qu'elle admet un maximum global unique par rapport à une valeur de p que l'on précisera. En déduire l'estimateur du maximum de vraisemblance du paramètre p noté $\hat{p}_{\text{MV}}$.
2. En utilisant l'expression de $E[X_i]$, donner un estimateur des moments de p noté $\hat{p}_{\text{Mo}}$.
3. Devant la difficulté d'étudier les performances des estimateurs $\hat{p}_{\text{MV}}$ et $\hat{p}_{\text{Mo}}$, on se propose d'estimer le paramètre $a = \frac{1}{p}$ à l'aide de de l'échantillon $(X_1, ..., X_n)$. Quel est l'estimateur du maximum de vraisemblance du paramètre a noté $\hat{a}_{\text{MV}}$? Cet estimateur est-il sans biais et convergent ?
4. Déterminer la borne de Cramér-Rao pour les estimateurs non-biaisés de a. L'estimateur $\hat{a}_{\text{MV}}$ est-il l'estimateur efficace du paramètre a ?

Exercice 3: Test statistique (4 points)

On considère une suite de variables aléatoires X_i indépendantes et de même loi loi géométrique à valeurs dans $\mathbb{N}^* = \{n \in \mathbb{N}, n \geq 1\}$ définie par

$$P[X = x_i] = p(1-p)^{x_i - 1}, \quad x_i = 1, 2, ...$$

On admettra qu'une telle loi est de moyenne $E[X_i] = \frac{1}{p}$ et de variance $\text{Var}(X_i) = \frac{q}{p^2}$ avec $q = 1 - p$. On désire tester les deux hypothèses

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p = p_1 \text{ (avec } \; p_1 > p_0) \end{cases}$$

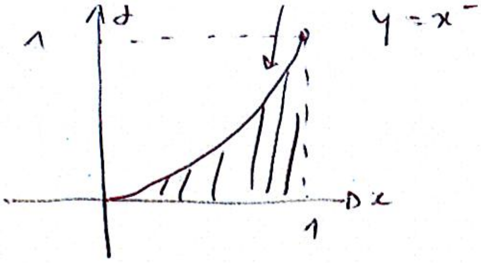
1. Vérifier que la statistique du test de Neyman-Pearson peut s'écrire $T = \sum_{i=1}^{n} X_i$ et déterminer la région critique associée.

2. On suppose que n est suffisamment grand pour pouvoir utiliser le théorème de la limite centrale.

   - Donner la loi approchée de T issue de ce théorème.
   - Déterminer les risques de première et de seconde espèce α et β en fonction du seuil du test noté K_α (et de n, p_0 et p_1) lorsqu'on confond la loi de T avec son approximation issue du théorème de la limite centrale.
   - En déduire les courbes COR de ce problème de détection. On posera

   $$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

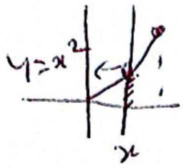
   et on notera $\Phi^{-1}(x)$ son inverse. Déterminer les paramètres qui influent sur la performance du test.



1) $\iint_{\mathbb{R}^2} f(x,y)\,dx\,dy = 1$ donc $\int_0^1 \left(\int_0^{x^2} A\,dy\right)dx = \int_0^1 A x^2\,dx = \frac{A}{3} = 1 \Rightarrow \boxed{A=3}$

1pt

2) $\underline{p(x,\cdot)} = ?$



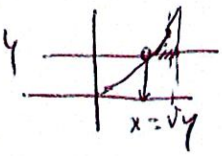
$$p(x,\cdot) = \begin{cases} 0 \text{ si } x \leq 0 \text{ ou } x \geq 1 \\ \int_0^{x^2} p(x,y)\,dy = \int_0^{x^2} 3\,dy = 3x^2 \end{cases}$$

$$\boxed{p(x,\cdot) = 3x^2\, \mathbb{1}_{[0,1]}(x)}$$ 1pt

On remarque que $\int_0^1 3x^2\,dx = \left[x^3\right]_0^1 = 1$

$\underline{p(\cdot,y)} = ?$

$$p(\cdot,y) = \begin{cases} 0 \text{ si } y \leq 0 \text{ ou } y \geq 1 \\ \int_{\sqrt{y}}^1 3\,dx = 3(1-\sqrt{y}) \quad y \in [0,1] \end{cases}$$



donc $\boxed{p(\cdot,y) = 3(1-\sqrt{y})\,\mathbb{1}_{[0,1]}(y)}$ 1pt

on remarque que $\int_{\mathbb{R}} p(\cdot,y)\,dy = 3\left[\int_0^1 dy - \int_0^1 \sqrt{y}\,dy\right] = 3\left\{[y]_0^1 - \left[\frac{y^{3/2}}{3/2}\right]_0^1\right\}$

$$= 3\left[1-\frac{2}{3}\right] = 1$$

On remarque que X suit une loi beta $B(3,1)$ donc $E[X] = \frac{a}{a+b} = \frac{3}{4}$ et

$\mathrm{Var}X = \frac{ab}{(a+b)^2(a+b+1)} = \frac{3}{16(5)} = \boxed{\frac{3}{80}}$ (a, b) 1pt

Pour Y, on a $E[Y] = \int_0^1 3y(1-\sqrt{y})\,dy = 3\left[\frac{y^2}{2}\right]_0^1 - 3\left[\frac{y^{5/2}}{5/2}\right]_0^1 = 3\left(\frac{1}{2}-\frac{2}{5}\right) = \frac{3}{10}$

$E[Y^2] = \int_0^1 3y^2(1-\sqrt{y})\,dy = 3\left[\frac{y^3}{3}\right]_0^1 - 3\left[\frac{y^{7/2}}{7/2}\right]_0^1 = 3\left(\frac{1}{3}-\frac{2}{7}\right) = \frac{1}{7}$

donc $\mathrm{Var}Y = E[Y^2] - E^2[Y] = \frac{1}{7} - \frac{9}{100} = \frac{37}{700}$ 1pt

Le domaine de définition du couple (X,Y) n'est pas un pavé donc les va X et Y sont <u>dépendantes</u>

1pt

La covariance du couple (X,Y) est $\mathrm{cov}(X,Y) = E[XY] - E[X]E[Y]$

On a $E[XY] = \iint xy\,p(x,y)\,dx\,dy = \int_0^1 \left[\int_0^{x^2} 4xy\,dy\right]dx = \int_0^1 A\frac{x^3}{2}\,dx = \frac{3}{8}$

La loi conditionnelle de $Y|X=x$ est définie par la densité:

$$p(y|x) = \frac{p(x,y)}{p(x,\cdot)} = \frac{3}{3x^2} \quad \text{si } x\in]0,1[\text{ et } 0<y<x^2<1$$

On en déduit, pour tout $x\in]0,1]$

$$\boxed{p(y|x) = \frac{1}{x^2}\mathbb{1}_D(y) \quad \text{où } D = \{y /\ 0<y<x^2\leq 1\} = \{y /\ y>0 \text{ et } y<x^2,\ y<1\}}$$

1pt

On remarque que $\int_0^{x^2} \frac{1}{x^2}\,dy = 1$

4) $\begin{cases} Z = \frac{Y}{X^2} \\ T = X \end{cases} \Longleftrightarrow \begin{cases} X = T \\ Y = ZT^2 \end{cases}$ avec

· Le changement de variables est donc bijectif de $D = \{(x,y)/\ 0\leq y\leq x^2\leq 1,\ x\geq 0\}$ dans Δ

Recherche de Δ

$$\begin{cases} 0\leq y \\ y\leq x^2 \\ x^2\leq 1 \\ x\geq 0 \end{cases} \Longleftrightarrow \begin{cases} zt^2\geq 0 \\ zt^2\leq t^2 \\ t^2\leq 1 \\ t>0 \end{cases} \Longleftrightarrow \begin{cases} z\geq 0 \\ z\leq 1 \\ -1\leq t\leq +1 \\ t>0 \end{cases} \Longleftrightarrow \begin{cases} t\in[0,1] \\ z\in[0,1] \end{cases}$$

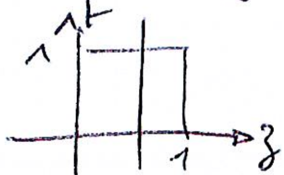
donc $\boxed{\Delta = [0,1]\times[0,1]}$

La densité du couple (Z,T) est $q(z,t) = A\times|J|$ avec $J = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial y}{\partial z} \\ \frac{\partial x}{\partial t} & \frac{\partial y}{\partial t} \end{vmatrix}$

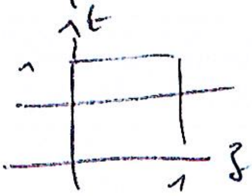
donc $J = \begin{vmatrix} 0 & t^2 \\ 1 & 2zt \end{vmatrix} = -t^2$

d'où $\boxed{q(z,t) = \begin{cases} 3t^2 & (z,t)\in\Delta \\ 0 & \text{sinon} \end{cases}}$ 2pts

Les lois marginales de Z et T sont



$$q(z,\cdot) = \begin{cases} 0 & \text{si } z\notin[0,1] \\ \int_0^1 3t^2\,dt = 1 & \text{si } z\in[0,1] \end{cases} \quad \text{i.e. } \boxed{Z\sim U([0,1])}$$



$$q(\cdot,t) = \begin{cases} 0 & \text{si } t\notin[0,1] \\ \int_0^1 3t^2\,dz & \text{si } t\in[0,1] \end{cases} \quad \text{i.e. } \boxed{q(\cdot,t) = 3t^2\mathbb{1}_{[0,1]}(t)}$$

1pt

On a $q(z,t) = q(z,\cdot) \times q(\cdot,t)$ $\forall (z,t)$ donc Z et T sont des va indépendantes (3)

Exercice 2

1) $L(x_1,\dots,x_n) = p^n \prod_{i=1}^{n} (1-p)^{x_i - 1} = \dfrac{p^n}{(1-p)^n} (1-p)^{\sum_{i=1}^n x_i}$

$\ln L(x_1,\dots,x_n) = n \ln p - n \ln(1-p) + \left(\sum_{i=1}^n x_i\right) \ln(1-p)$

$\dfrac{\partial \ln}{\partial p} \geq 0 \iff \dfrac{n}{p} - n\left(\dfrac{-1}{1-p}\right) + \left(\sum_{i=1}^n x_i\right)\left(\dfrac{-1}{1-p}\right) \geq 0$

$\iff \dfrac{n}{p(1-p)} - \dfrac{\sum x_i}{1-p} \geq 0 \iff \dfrac{n}{p} \geq \sum x_i$

$\iff p \leq \dfrac{n}{\sum x_i}$

donc

p	0		$\frac{n}{\sum x_i}$		1
$\frac{\partial \ln L}{\partial p}$		$+$	0	$-$	
$\ln L$		$\nearrow$		$\searrow$	

0,5pt

donc $\ln L$ admet un maximum global unique atteint en $p = \dfrac{n}{\sum x_i}$

On en déduit $\boxed{\hat{p}_{MV} = \dfrac{n}{\sum_{i=1}^n X_i}}$ 0,5pt

2) $E[X_i] = \dfrac{1}{p} \Rightarrow \boxed{\hat{p}_{MO} = \dfrac{1}{\frac{1}{n}\sum X_i} = \hat{p}_{MV}}$ 1pt

3) $a = \dfrac{1}{p}$ donc d'après la propriété d'invariance fonctionnelle

$\hat{a}_{MV} = \left(\hat{p}_{MV}\right)^{-1} = \dfrac{1}{n} \sum_{i=1}^n X_i$ 1pt

$E[\hat{a}_{MV}] = E(X_i) = \dfrac{1}{p} = a$ donc $\hat{a}_{MV}$ estimateur non biaisé de a 1pt

$Var\left[\hat{a}_{MV}\right] = \dfrac{1}{n} Var(X_1) = \dfrac{q}{p^2 n} = \dfrac{1}{n} \dfrac{1-\frac{1}{a}}{\frac{1}{a^2}} = \dfrac{a^2 - a}{n} = \boxed{\dfrac{a(a-1)}{n}}$

4) $\dfrac{\partial^2 \ln L}{\partial a^2}$? $\quad \ln L = -n\ln a - n\ln(1-\frac{1}{a}) + (\sum x_i)\ln(1-\frac{1}{a})$ (4)

$$= -n\cancel{\ln a} - n\ln(a-1) + n\cancel{\ln a} + (\sum x_i)\ln(1-\frac{1}{a})$$

$$\frac{\partial \ln L}{\partial a} = \frac{-n}{a-1} + \sum x_i \frac{\frac{1}{a^2}}{1-\frac{1}{a}} = \frac{n}{1-a} + \sum x_i \frac{1}{\underbrace{a(a-1)}_{a^2-a}}$$

$$\frac{\partial^2 \ln L}{\partial a^2} = \frac{n}{(a-1)^2} + (\sum x_i)\frac{-(2a-1)}{a^2(a-1)^2}$$

$$E\left[\frac{\partial^2 \ln L}{\partial a^2}\right] = \frac{n}{(a-1)^2} + \frac{1-2a}{a^2(a-1)^2} \times na$$

$$= \frac{n[a^2+a-2a^2)}{a^2(a-1)^2} = \frac{n\,a(1-a)}{a^2(1-a)^2} = \frac{n}{a(1-a)}$$

donc la borne de Cramer-Rao est $\boxed{BCR = \frac{a(a-1)}{n}}$ 2pts

Ex3

1) Rejet de H_0 si $\dfrac{(1-p_1)^{\sum x_i}}{(1-p_0)^{\sum x_i}} > S_\alpha \iff (\sum x_i)\ln\left(\frac{1-p_1}{1-p_0}\right) > k_\alpha$

$p_1 > p_0$ donc $1-p_1 < 1-p_0$ donc $\dfrac{1-p_1}{1-p_0} < 1$ donc

1pt

$\boxed{\text{Rejet de } H_0 \text{ si } \sum x_i < S_\alpha}$

2) • $T \simeq N\left(\frac{n}{p}, \frac{nq}{p^2}\right)$ 1pt

• $\alpha = P\left[\sum X_i < S_\alpha \mid \sum X_i \sim N\left(\frac{n}{p_0}, \frac{nq_0}{p_0^2}\right)\right]$

$$= P\left[\underbrace{\frac{\sum X_i - \frac{n}{p_0}}{\sqrt{\frac{nq_0}{p_0^2}}}}_{U} < \frac{S_\alpha - \frac{n}{p_0}}{\sqrt{\frac{nq_0}{p_0^2}}} \;\middle|\; U \sim N(0,1)\right] = \boxed{\phi\left(\frac{S_\alpha - \frac{n}{p_0}}{\sqrt{\frac{nq_0}{p_0^2}}}\right) = \alpha}$$

Ex1

1) $A = 3$ (1pt) — 1

2) $p(x,\cdot) = 3x^2 \mathbb{1}_{[0,1]}(x)$ (1pt)

$p(\cdot,y) = 3(1-\sqrt{y})\mathbb{1}_{[0,1]}(y)$ (1pt)

$X \sim B(3,1)$ $E[X] = \frac{3}{4}$ $Var[X] = \frac{3}{80}$ ~~(1pt)~~ (0,5)

$E[Y] = 3\left(\frac{1}{2} - \frac{2}{5}\right) = \frac{3}{10}$

$Var[Y] = \frac{~~37~~\,91}{700} = \frac{1}{7} - \frac{9}{100}$ (1pt)

Non Indépendance de X et de Y ~~(1pt)~~ (0,5)

~~5~~ 4

3) $p(y|x) = \frac{1}{x^2}$ $1 > y > 0$, $y < x^2$ (1pt) — 1

4) $\Delta = [0,1) \times [0,1]$ (1pt)

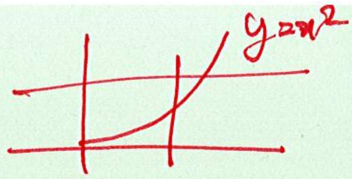
$J = -t^2$

$q(z,t) = 3t^2 \mathbb{1}_{\Delta}(z,t)$ (1pt)

$Z \sim U([0,1])$ (1pt)

$q(\cdot,t) = 3t^2 \mathbb{1}_{[0,1]}(t)$ (1pt)

4



$$\int_0^1 \left(\int_0^{x^2} A\,dy \right) dx$$

$$\int_0^1 \left[\int_{\sqrt{y}}^1 A\,dx \right] dy$$

Ex2

$\hat{p}_{MV} = \frac{1}{\bar{X}}$ (1) ~~(0,5)~~ Tableau de variation (1) ~~(0,5)~~ — 2

$\hat{p}_{MO} = \frac{1}{\widehat{E[X_i]}} = \frac{1}{\bar{X}}$ — 1

$\hat{a}_M = \bar{X}$ — 0,5

$E[\hat{a}_{MV}] = a$ $Var\,\hat{a} = \frac{a(a-1)}{n}$ — 1,5 — 2

$BCR = \frac{a(a-1)}{n}$ — 1,5

estimateur efficace — 0,5 — 2

Rejet de H_0 si $\sum x_i < s_\alpha$ — 1pt

$T \sim N\left(\frac{n}{p}, \frac{nq}{p^2}\right)$ — 1pt

$\alpha = \phi\left(\frac{s_\alpha - \frac{n}{p_0}}{\sqrt{\frac{nq_0}{p_0^2}}}\right)$ — 1pt

1pt

$\pi = 1 - \beta$

$= \phi\left(\frac{\phi^{-1}(\alpha)\sqrt{\frac{nq_0}{p_0^2}} + \frac{n}{p_0} - \frac{n}{p_1}}{\sqrt{\frac{nq_1}{p_1^2}}}\right)$

Analyse — 1pt