
EXAMEN STATISTIQUE - 1SN

Lundi 15 Janvier 2018

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

Exercice 1 : Estimation (9 points)

On considère n variables aléatoires X_1, \dots, X_n indépendantes suivant la même loi discrète à valeurs dans l'ensemble des entiers non nuls $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$ définie par

$$P[X_i = x_i; \theta] = \theta(1 - \theta)^{x_i - 1}, \quad x_i \in \mathbb{N}^* = \{1, 2, \dots\}$$

avec $\theta \in]0, 1[$. On admettra que la moyenne et la variance d'une telle loi appelée "loi géométrique de paramètre θ " sont définies par $E[X_i] = \frac{1}{\theta}$ et $\text{var}(X_i) = \frac{1-\theta}{\theta^2}$.

1. (2pts) Montrer que la vraisemblance de (x_1, \dots, x_n) admet un unique maximum global pour une valeur de θ que l'on déterminera. En déduire l'estimateur du maximum de vraisemblance du paramètre θ noté $\hat{\theta}_{MV}$.
2. (2pts) Rappeler les propriétés asymptotiques de l'estimateur $\hat{\theta}_{MV}$. Devant la difficulté d'étudier ces propriétés pour n fini, on propose d'estimer le paramètre $a = \frac{1}{\theta}$. Quel est l'estimateur du maximum de vraisemblance du paramètre a noté \hat{a}_{MV} ? Cet estimateur est-il sans biais et convergent?
3. (2pts) Déterminer la borne de Cramer-Rao pour un estimateur non biaisé du paramètre a . L'estimateur \hat{a}_{MV} est-il l'estimateur efficace du paramètre a ?
4. (3pts) On désire maintenant construire un estimateur Bayésien du paramètre θ . Puisque ce paramètre vérifie la contrainte $\theta \in]0, 1[$, il est naturel de définir une loi a priori (appelée loi beta de paramètres α et β) de densité

$$p(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)} I_{]0,1[}(\theta)$$

où $I_{]0,1[}(\theta)$ est la fonction indicatrice sur l'intervalle $]0, 1[$ ($I_{]0,1[}(\theta) = 1$ si $\theta \in]0, 1[$ et $I_{]0,1[}(\theta) = 0$ si $\theta \notin]0, 1[$) et où $B(\alpha, \beta)$ est la fonction beta dont l'expression n'est pas importante dans cet exercice.

- Montrer que la loi a posteriori de $\theta | x_1, \dots, x_n$ est aussi une loi beta dont on précisera les paramètres.
- Déterminer l'estimateur du maximum a posteriori du paramètre θ noté $\hat{\theta}_{MAP}$ et étudier son comportement lorsque $n \rightarrow \infty$.

Exercice 2 : Test Statistique (7 points)

On considère n variables aléatoires X_1, \dots, X_n indépendantes suivant la même loi géométrique à valeurs dans l'ensemble des entiers non nuls $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$ définie par

$$P[X_i = x_i; \theta] = \theta(1 - \theta)^{x_i - 1}, \quad x_i \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}.$$

avec $\theta \in]0, 1[$. On rappelle que la moyenne et la variance d'une loi géométrique de paramètre θ sont définies par $E[X_i] = \frac{1}{\theta}$ et $\text{var}(X_i) = \frac{1-\theta}{\theta^2}$. On considère le test d'hypothèses simples

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1.$$

- (2pts) Déterminer la statistique du test de Neyman Pearson notée T_n et indiquer la région critique de ce test pour $\theta_1 > \theta_0$ et $\theta_1 < \theta_0$. Dans la suite de cet exercice, on supposera $\theta_1 > \theta_0$. La décision prise à l'aide du test de Neyman Pearson est-elle en accord avec la moyenne d'une loi géométrique définie par $E[X_i] = \frac{1}{\theta}$?
- (1pt) Déterminer la loi approchée de la statistique T_n résultant de l'application du théorème de la limite centrale.
- (1pt) On note

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

la fonction de répartition d'une loi normale $\mathcal{N}(0, 1)$ et F^{-1} son inverse. Déterminer la valeur du seuil du test de Neyman Pearson notée K_α en fonction de n, θ_0 , du risque de première espèce α et de F^{-1} .

- (1pt) Déterminer la puissance du test en fonction du seuil K_α , de n, θ_1 et de F .
- (2pts) Déterminer les courbes COR du test étudié dans cet exercice et tracer la forme de ces courbes pour différentes valeurs de n .

Exercice 3 : Test d'adéquation (4 points)

On observe 20 réalisations d'une variable aléatoire discrète regroupées dans le tableau ci-dessous et on se pose la question de savoir si ces observations proviennent d'une loi géométrique de paramètre $\theta = 0.4$ (voir définition à l'exercice précédent). Pour cela, on effectue un test du χ^2 avec les 4 classes $C_1 = \{1\}$, $C_2 = \{2\}$, $C_3 = \{3\}$ et $C_4 = \{4, \dots\}$

$x_1 = 1$	$x_2 = 1$	$x_3 = 3$	$x_4 = 2$	$x_5 = 2$	$x_6 = 6$	$x_7 = 3$	$x_8 = 5$	$x_9 = 4$	$x_{10} = 3$
$x_{11} = 2$	$x_{12} = 6$	$x_{13} = 1$	$x_{14} = 1$	$x_{15} = 2$	$x_{16} = 2$	$x_{17} = 3$	$x_{18} = 1$	$x_{19} = 2$	$x_{20} = 5$

- (1pt) Déterminer les probabilités des différentes classes notées $p_i, i = 1, \dots, 4$ en fonction de θ . On admettra que les valeurs numériques de ces probabilités pour $\theta = 0.4$ sont $p_1 = 0.4, p_2 = 0.24, p_3 = 0.144$ et $p_4 = 0.216$.
- (1pt) En déduire la valeur de la statistique du test du χ^2 notée ϕ_n (on écrira l'expression de ϕ_n sous la forme d'une somme pondérée de carrés qu'on ne cherchera pas à calculer).
- (1pt) Quelle est la loi de ϕ_n lorsque les données x_i sont issues d'une loi géométrique de paramètre $\theta = 0.4$?
- (1pt) Pour effectuer le test du χ^2 , on doit comparer la valeur de ϕ_n à un seuil. Déterminer ce seuil en fonction de la fonction de répartition inverse d'une loi du χ^2 et du risque de première espèce du test noté α .

LOIS DE PROBABILITÉ CONTINUES

m : moyenne σ^2 : variance F. C. : fonction caractéristique

LOI	Densité de probabilité	m	σ^2	F. C.
Uniforme	$f(x) = \frac{1}{b-a}$ $x \in]a, b[$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{itb} - e^{ita}}{it(b-a)}$
Gamma $\Gamma(\theta, \nu)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\theta x} x^{\nu-1}$ $\theta > 0, \nu > 0$ $x \geq 0$	$\frac{\nu}{\theta}$	$\frac{\nu}{\theta^2}$	$\frac{1}{(1 - i\frac{t}{\theta})^\nu}$
Inverse gamma IG(θ, ν)	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\frac{\theta}{x}} \frac{1}{x^{\nu+1}}$ $\theta > 0, \nu > 0$ $x \geq 0$	$\frac{\theta}{\nu-1}$ si $\nu > 1$	$\frac{\theta^2}{(\nu-1)^2(\nu-2)}$ si $\nu > 2$	(*)
Première loi de Laplace	$f(x) = \frac{1}{2} e^{- x }$	0	2	$\frac{1}{1+t^2}$
Normale $\mathcal{N}(m, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2	$e^{imt - \frac{\sigma^2 t^2}{2}}$
Khi ₂ χ_ν^2 $\Gamma(\frac{1}{2}, \frac{\nu}{2})$	$f(x) = k e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}$ $k = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})}$ $\nu \in \mathbb{N}^*, x \geq 0$	ν	2ν	$\frac{1}{(1-2it)^{\frac{\nu}{2}}}$
Cauchy $c_{\lambda, \alpha}$	$f(x) = \frac{1}{\pi\lambda \left(1 + \left(\frac{x-\alpha}{\lambda}\right)^2\right)}$ $\lambda > 0, \alpha \in \mathbb{R}$	(-)	(-)	$e^{i\alpha t - \lambda t }$
Beta	$f(x) = k x^{a-1} (1-x)^{b-1}$ $k = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ $a > 0, b > 0, x \in]0, 1[$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	(*)

LOIS DE PROBABILITÉ DISCRÈTES

m : moyenne σ^2 : variance **F. C.** : fonction caractéristique

$p_k = P[X = k]$ $p_{1,\dots,m} = P[X_1 = k_1, \dots, X_m = k_m]$

LOI	Probabilités	m	σ^2	F. C.
Uniforme	$p_k = \frac{1}{n}$ $k \in \{1, \dots, n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\frac{e^{it}(1 - e^{itn})}{n(1 - e^{it})}$
Bernoulli	$p_1 = P[X = 1] = p$ $p_0 = P[X = 0] = q$ $p \in [0, 1]$ $q = 1 - p$	p	pq	$pe^{it} + q$
Binomiale $B(n, p)$	$p_k = C_n^k p^k q^{n-k}$ $p \in [0, 1]$ $q = 1 - p$ $k \in \{0, 1, \dots, n\}$	np	npq	$(pe^{it} + q)^n$
Binomiale négative	$p_k = C_{n+k-1}^{n-1} p^n q^k$ $p \in [0, 1]$ $q = 1 - p$ $k \in \mathbb{N}$	$n \frac{q}{p}$	$n \frac{q}{p^2}$	$\left(\frac{p}{1 - qe^{it}}\right)^n$
Multinomiale	$p_{1,\dots,m} = \frac{n!}{k_1! \dots k_m!} p_1^{k_1} \dots p_m^{k_m}$ $p_j \in [0, 1]$ $q_j = 1 - p_j$ $k_j \in \{0, 1, \dots, n\}$ $\sum_{j=1}^m k_j = n$ $\sum_{j=1}^m p_j = 1$	np_j	Variance : $np_j q_j$ Covariance : $-np_j p_k$	$\left(\sum_{j=1}^m p_j e^{it}\right)^n$
Poisson $P(\lambda)$	$p_k = e^{-\lambda} \frac{\lambda^k}{k!}$ $\lambda > 0$ $k \in \mathbb{N}$	λ	λ	$\exp[\lambda(e^{it} - 1)]$
Géométrique	$p_k = pq^{k-1}$ $p \in [0, 1]$ $q = 1 - p$ $k \in \mathbb{N}^*$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^{it}}{1 - qe^{it}}$