
EXAMEN STATISTIQUE - 1MF2E

Lundi 25 Mars 2024 (8h-9h45)

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

Exercice 1 : Estimation (10 points)

On considère n observations x_1, \dots, x_n issues d'un vecteur (X_1, \dots, X_n) de n variables aléatoires indépendantes de même loi binomiale négative définie par

$$P[X_i = x_i; p] = \binom{k + x_i - 1}{k - 1} p^k q^{x_i}, \quad x_i \in \mathbb{N}$$

avec $p \in]0, 1[$, $q = 1 - p$, $k \in \mathbb{N}$ et $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. On fera attention car les notations sont légèrement différentes de celles des tables puisqu'ici n est le nombre d'observations et les paramètres de cette loi sont notés p et k . En tenant compte de ce changement de notation, cette loi est de moyenne $E[X_i] = k \frac{q}{p}$, de variance $\text{var}(X_i) = k \frac{q}{p^2}$ et de fonction caractéristique $\left(\frac{p}{1 - qe^{it}} \right)^k$. On supposera dans cet exercice que k est un paramètre connu et que p est un paramètre inconnu que l'on cherche à estimer.

1. (2pts) Déterminer l'estimateur du vraisemblance du paramètre p noté \hat{p}_{MV} . Comme cet estimateur semble difficile à étudier, on s'intéresse maintenant au paramètre $\theta = \frac{1-p}{p}$. Montrer (en utilisant par exemple l'invariance fonctionnelle) que l'estimateur du maximum de vraisemblance de θ est

$$\hat{\theta}_{MV} = \frac{1}{nk} \sum_{i=1}^n X_i.$$

2. (2pt) L'estimateur $\hat{\theta}_{MV}$ est-il un estimateur sans biais et convergent du paramètre θ ?
3. (2pt) Déterminer la borne de Cramer-Rao pour un estimateur non biaisé du paramètre θ . L'estimateur $\hat{\theta}_{MV}$ est-il l'estimateur efficace du paramètre θ ?
4. (1pt) Déterminer un estimateur des moments du paramètre θ noté $\hat{\theta}_{Mo}$.
5. (3pts) On suppose désormais que le paramètre θ est muni d'une loi a priori de densité

$$f(\theta) = \frac{1}{(1 + \theta)^2} I_{\mathbb{R}^+}(\theta).$$

où $I_{\mathbb{R}^+}$ est la fonction indicatrice sur \mathbb{R}^+ ($I_{\mathbb{R}^+}(\theta) = 1$ si $\theta > 0$ et 0 sinon).

- Expliquer les connaissances sur θ apportées par la forme de $f(\theta)$.
- Déterminer l'estimateur du maximum a posteriori du paramètre θ noté $\hat{\theta}_{MAP}$ et vérifier que $\hat{\theta}_{MAP} < \hat{\theta}_{MV}$. Est ce que ceci est cohérent avec l'expression de la densité a priori $f(\theta)$?

Exercice 2 : Tests Statistiques (10 points)

On considère n observations x_1, \dots, x_n issues d'un vecteur (X_1, \dots, X_n) de n variables aléatoires indépendantes de même loi binomiale négative définie par

$$P[X_i = x_i; p] = \binom{k + x_i - 1}{k - 1} p^k q^{x_i}, \quad x_i \in \mathbb{N}$$

avec $p \in]0, 1[$, $q = 1 - p$, $k \in \mathbb{N}$ et $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. On fera attention car les notations sont légèrement différentes de celles des tables puisqu'ici n est le nombre d'observations et les paramètres de cette loi sont notés p et k . En tenant compte de ce changement de notation, cette loi est de moyenne $E[X_i] = k \frac{q}{p}$, de variance $\text{var}(X_i) = k \frac{q}{p^2}$ et de fonction caractéristique $\left(\frac{p}{1 - qe^{it}} \right)^k$. On supposera dans cet exercice que k est un paramètre connu et on désire utiliser les observations x_1, \dots, x_n pour déterminer si $p = p_0 > 0$ ou si $p = p_1 > p_0$. On considère donc le test d'hypothèses

$$H_0 : p = p_0, \quad H_1 : p = p_1 \quad \text{avec } p_1 > p_0 > 0.$$

1. (2pt) Déterminer la statistique T_n du test de Neyman Pearson et la région critique associée.
2. (1pt) Exprimer la moyenne et la variance de T_n en fonction de k , n et $\theta = \frac{q}{p} = \frac{1-p}{p}$. En déduire la loi approchée de T_n sous les deux hypothèses H_0 et H_1 provenant du théorème central limite (on exprimera les paramètres de cette loi en fonction de k , n et θ).
3. (2pts) On note F la fonction de répartition d'une loi du normale $\mathcal{N}(0, 1)$. En utilisant la loi approchée de T_n déterminée à la question précédente, exprimer le risque de première espèce α en fonction du seuil du test de Neyman Pearson noté S_α , de $F(\alpha)$, nk et θ_0 . En déduire la valeur du seuil S_α en fonction de $F^{-1}(\alpha)$, nk et θ_0 .
4. (2pts) En utilisant la loi approchée de T_n déterminée à la question 2), déterminer les caractéristiques opérationnelles du récepteur (courbes COR) pour ce test et montrer qu'elles ne dépendent que de \sqrt{nk} , $F^{-1}(\alpha)$ et de deux fonctions dépendant des paramètres θ_0 et θ_1 . Analyser les performances du test en fonction de nk et tracer l'allure approximative des courbes COR pour différentes valeurs de ce paramètre.
5. (3pts) On désire vérifier que les observations suivent une loi binomiale négative à l'aide d'un test du χ^2 . Le tableau suivant résume les quantités nécessaires pour effectuer le test

x_i	0	1	2	3	4	5	6	7
n_k	70	38	17	10	9	3	2	1
$n\pi_k$	69.49	37.60	20.10	10.70	5.69	3.02	1.60	0.85

où n_k est le nombre d'observations égales à k avec $k \in \{0, \dots, 7\}$, $n = \sum_{k=0}^7 n_k = 150$ est le nombre total d'observations et $\pi_k = P[X_i = k]$ a été calculée avec une loi binomiale négative dont les deux paramètres k et p ont été estimés avec la méthode du maximum de vraisemblance. Les trois dernières observations sont regroupées dans une même classe, de manière à obtenir 6 classes définies par $C_0 = \{0\}$, $C_1 = \{1\}$, $C_2 = \{2\}$, $C_3 = \{3\}$, $C_4 = \{4\}$ et $C_5 = \{5, 6, 7\}$.

- Expliquer pourquoi on a regroupé les valeurs $\{5\}$, $\{6\}$ et $\{7\}$ dans une même classe.
- Exprimer la statistique du test du χ^2 en fonction des données du problème pour les 6 classes définies ci-dessus (sans chercher à la calculer).
- Rappeler la région critique du test du χ^2 .
- Quelle loi doit être utilisée pour calculer le seuil de ce test noté S_α ?
- Donner l'expression de S_α en fonction de α et de l'inverse de la fonction de répartition de la loi trouvée à la question précédente.
- Si on augmente la valeur de α , la valeur de S_α augmente-t-elle ou diminue-t-elle ?

LOIS DE PROBABILITÉ DISCRÈTES

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

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

LOIS DE PROBABILITÉ CONTINUES

LOI	Densité de probabilité	Moyenne	Variance	Fonction Caractéristique
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 $\mathcal{G}(\nu, \theta)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\theta x} x^{\nu-1}$ $\theta > 0, \nu > 0$ $x \geq 0$ avec $\Gamma(n+1) = n! \forall n \in \mathbb{N}$	$\frac{\nu}{\theta}$	$\frac{\nu}{\theta^2}$	$\frac{1}{(1 - i\frac{t}{\theta})^\nu}$
Inverse gamma $\mathcal{IG}(\nu, \theta)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\frac{\theta}{x}} \frac{1}{x^{\nu+1}}$ $\theta > 0, \nu > 0$ $x \geq 0$ avec $\Gamma(n+1) = n! \forall n \in \mathbb{N}$	$\frac{\theta}{\nu-1}$ si $\nu > 1$	$\frac{\theta^2}{(\nu-1)^2(\nu-2)}$ si $\nu > 2$	(*)
Première loi de Laplace	$f(x) = \frac{1}{2} e^{- x }, \quad x \in \mathbb{R}$	0	2	$\frac{1}{1+t^2}$
Normale univariée $\mathcal{N}(m, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$	m	σ^2	$e^{imt - \frac{\sigma^2 t^2}{2}}$
Normale multivariée $\mathcal{N}_p(\mathbf{m}, \Sigma)$	$f(x) = K e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \Sigma^{-1}(\mathbf{x}-\mathbf{m})}$ $K = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}}$ $x \in \mathbb{R}^p$	\mathbf{m}	Σ	$e^{i\mathbf{u}^T \mathbf{m} - \frac{1}{2} \mathbf{u}^T \Sigma \mathbf{u}}$
Khi2 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 $B(a, b)$	$f(x) = k x^{a-1} (1-x)^{b-1}$ $k = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ $a > 0, b > 0$ $x \in]0, 1[$ avec $\Gamma(n+1) = n! \forall n \in \mathbb{N}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	(*)