

---

EXAMEN STATISTIQUE - 1MF2E

Lundi 4 Avril 2022 - 8h00-9h45

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

---

**Exercice 1 : Estimation (11 points)**

Un ensemble de  $n$  votants s'exprime pour 3 candidats dont deux d'entre eux ont la même probabilité  $\theta$  d'être élus. La vraisemblance associée à ce modèle statistique est définie par une loi multinomiale d'expression  $L(n_1, n_2, n_3; \theta) = P[N_1 = n_1, N_2 = n_2, N_3 = n_3] = \frac{n!}{n_1!n_2!n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}$  où  $N_i$  est la variable aléatoire associée au nombre de voix obtenues pour le  $i$ ème candidat (avec  $n_1 + n_2 + n_3 = n$ ) et  $\mathbf{p} = (p_1, p_2, p_3) = (\theta, \theta, 1 - 2\theta)$  regroupe les probabilités que chaque candidat reçoive une voix, qui dépend d'un unique paramètre inconnu  $\theta \in ]0, 1/2[$ . On admettra que pour un tel modèle statistique, la loi marginale de  $N_i$  est une loi binomiale  $B(n, p_i)$  de moyenne  $E[N_i] = np_i$  et de variance  $\text{var}(N_i) = np_i(1 - p_i)$ .

1. (2pts) Déterminer l'estimateur du maximum de vraisemblance du paramètre  $\theta$  noté  $\hat{\theta}_{MV}$  en utilisant la vraisemblance  $L(n_1, n_2, n_3; \theta)$  et montrer qu'il ne dépend que de  $n$  et de  $N_3$  (on prendra soin de vérifier que cette vraisemblance admet un unique maximum global).
2. (2pts) Déterminer la moyenne et la variance de l'estimateur  $\hat{\theta}_{MV}$ . Est-il un estimateur sans biais et convergent du paramètre  $\theta$  ?
3. (2pts) L'estimateur  $\hat{\theta}_{MV}$  est-il l'estimateur efficace du paramètre  $\theta$  ?
4. (5pts) On suppose maintenant qu'on dispose d'une information a priori sur le paramètre  $\theta$  résumée dans la loi a priori  $p(\theta)$  définie par une loi Beta à trois paramètres notée  $\mathcal{B}(\alpha, \beta, \gamma)$  de densité

$$\pi(\theta) = \frac{\theta^{\alpha-1}(\gamma - \theta)^{\beta-1}}{\gamma^{\alpha+\beta-1} \mathbf{B}(\alpha, \beta)} \mathcal{I}_{]0, \gamma[}(\theta)$$

avec  $\gamma = \frac{1}{2}$ ,  $\mathbf{B}(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$  et où  $\mathcal{I}_{]0, \gamma[}$  est la fonction indicatrice sur  $]0, \gamma[$  ( $\mathcal{I}_{]0, \gamma[}(x) = 1$  si  $x \in ]0, \gamma[$  et  $\mathcal{I}_{]0, \gamma[}(x) = 0$  sinon). On admettra qu'une telle loi est de moyenne  $\frac{\alpha\gamma}{\alpha+\beta}$  et de variance  $\frac{\gamma^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta-1)}$ .

- La loi a priori de  $\theta$  est de support  $]0, 1/2[$ . Pourquoi ne peut-on pas avoir  $\theta > 1/2$  ?
- Déterminer l'estimateur MAP de  $\theta$  obtenu à l'aide de la vraisemblance  $L(n_1, n_2, n_3; \theta)$  et de la loi a priori  $\pi(\theta)$  noté  $\hat{\theta}_{MAP}$ .
- Expliquer le comportement de  $\hat{\theta}_{MAP}$  lorsque  $n_1, n_2$  et  $n_3$  tendent vers  $+\infty$ .
- Montrer que la loi a posteriori de  $\theta|n_1, n_2, n_3$  est une loi beta à trois paramètres (on déterminera ces trois paramètres). En déduire l'estimateur de la moyenne a posteriori du paramètre  $\theta$  noté  $\hat{\theta}_{MMSE}$ .

## Exercice 2 : Test Statistique (10 points)

On considère un triplet de variables aléatoires  $(N_1, N_2, N_3)$  de loi multinomiale définie par  $P[N_1 = n_1, N_2 = n_2, N_3 = n_3] = \frac{n!}{n_1!n_2!n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}$  avec  $n_1 + n_2 + n_3 = n$  et  $\mathbf{p} = (p_1, p_2, p_3) = (\theta, \theta, 1 - 2\theta)$ . On admettra que pour un tel modèle statistique, la loi marginale de  $N_i$  est une loi binomiale  $B(n, p_i)$  de moyenne  $E[N_i] = np_i$  et de variance  $\text{var}(N_i) = np_i(1 - p_i)$ . On veut à l'aide de l'observation de  $(N_1, N_2, N_3)$  tester les deux hypothèses suivantes

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1 \quad \text{avec } \theta_1 > \theta_0.$$

1. (2pts) À l'aide du théorème de Neyman-Pearson, montrer que la statistique du test le plus puissant est  $T_n = N_3$  (on n'oubliera pas que  $N_1 + N_2 + N_3 = n$ ) et indiquer la région critique de ce test.
2. (1pt) En utilisant le théorème central limite, déterminer la loi asymptotique de  $T_n$  sous les deux hypothèses.
3. (2pts) En utilisant la loi asymptotique de la question précédente, déterminer les risques de première et seconde espèce  $\alpha$  et  $\beta$  du test en fonction des paramètres  $\theta_0$  et  $\theta_1$  et de la fonction de répartition d'une loi normale  $\mathcal{N}(0, 1)$  notée  $F$ .
4. (2pts) Déterminer les courbes COR associées à ce test et tracer l'allure de ces courbes pour différentes valeurs de  $n$ .
5. (3pts) On désire à partir d'un ensemble de  $K = 10$  observations de  $N_1$  notées  $(n_{11}, \dots, n_{1K})$  et données dans le tableau ci-dessous déterminer s'il est raisonnable de supposer que  $N_1$  suit une loi binomiale  $B(K, \theta)$ .

2	3	3	2	2	0	4	3	1	0
---	---	---	---	---	---	---	---	---	---

- Expliquer pourquoi on est amené à choisir  $\theta = 0.2$  pour effectuer ce test.
- Les probabilités associées à une loi binomiale  $B(K, \theta)$  avec  $\theta = 0.2$  et  $K = 10$  sont données dans le tableau ci-dessous

$p_0$	$p_1$	$p_2$	$p_3$	$p_4$	$\sum_{k=5}^{10} p_k$
0.11	0.27	0.30	0.20	0.09	0.03

On décide de faire un test du  $\chi^2$  avec 3 classes notées  $C_1, C_2$  et  $C_3$ . Justifier le choix suivant pour ces trois classes :  $C_1 = \{0, 1\}$ ,  $C_2 = \{2\}$  et  $C_3 = \{3, \dots, 10\}$ .

- Exprimer la statistique de ce test notée  $\phi$  en fonction des données du problème (sans chercher à la calculer) et préciser la région critique.
- Quelle est la loi asymptotique de  $\phi$  ?
- Les seuils obtenus pour  $\alpha = 0.05$  et  $\alpha = 0.01$  sont  $S_{0.05} = 5.99$  et  $S_{0.01} = 9.21$ . Expliquer pourquoi  $S_{0.05} < S_{0.01}$ .

## LOIS DE PROBABILITÉ CONTINUES

m : moyenne     $\sigma^2$  : variance    F. C. : fonction caractéristique

LOI	Densité de probabilité	m	$\sigma^2$	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 $\mathcal{G}(\nu, \theta)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\theta x} x^{\nu-1}$ $\theta > 0, \nu > 0$ $x \geq 0$ 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	$\sigma^2$	$e^{imt - \frac{\sigma^2 t^2}{2}}$
Normale multivariée $\mathcal{N}_p(\mathbf{m}, \Sigma)$	$f(x) = K e^{-\frac{1}{2}(x-m)^T \Sigma^{-1} (x-m)}$ $K = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}}$ $x \in \mathbb{R}^p$	$\mathbf{m}$	$\Sigma$	$e^{i\mathbf{u}^T \mathbf{m} - \frac{1}{2} \mathbf{u}^T \Sigma \mathbf{u}}$
Khi2 $\chi_\nu^2$ $\Gamma(\frac{1}{2}, \frac{\nu}{2})$	$f(x) = k e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}$ $k = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})}$ $\nu \in \mathbb{N}^*, x \geq 0$	$\nu$	$2\nu$	$\frac{1}{(1-2it)^{\frac{\nu}{2}}}$
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 $\Gamma(n+1) = n! \forall n \in \mathbb{N}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	(*)

## LOIS DE PROBABILITÉ DISCRÈTES

**m** : moyenne     $\sigma^2$  : 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	$\sigma^2$	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 $P(\lambda)$	$p_k = e^{-\lambda} \frac{\lambda^k}{k!}$ $\lambda > 0$ $k \in \mathbb{N}$	$\lambda$	$\lambda$	$\exp[\lambda(e^{it} - 1)]$
Géométrique	$p_k = pq^{k-1}$ $p \in [0, 1]$ $q = 1 - p$ $k \in \mathbb{N}^*$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^{it}}{1 - qe^{it}}$